Understanding and Preventing Disease-Associated Anaerobic Choline Metabolism by the Human Gut Microbiota

Understanding and Preventing Disease-Associated Anaerobic Choline Metabolism by the Human Gut Microbiota

A dissertation presented

by

Smaranda Bodea

to

The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

April 2016

Dissertation advisor: Professor Emily P. Balskus Smaranda Bodea

Understanding and Preventing Disease-Associated Anaerobic Choline Metabolism by the Human Gut Microbiota

Abstract

The consortium of microorganisms inhabiting the human gastrointestinal tract, our gut microbiota, has an extensive set of metabolic capabilities that directly influence human health. Over the past decade, DNA sequencing has significantly improved our knowledge of this microbial community's composition. However, the molecular details of how gut bacterial metabolism impacts human health and disease are largely unknown. We propose that by combining chemical knowledge with bioinformatics, we can uncover microbial metabolic pathways that contribute to human disease. By developing small molecule inhibitors specifically targeting these pathways, we can further elucidate the roles of bacterial metabolism in disease and create a new paradigm for therapeutics.

To illustrate this strategy's potential, we are investigating the catabolism of the essential nutrient choline by anaerobic gut microbes. Within the human gastrointestinal tract, bacteria can process choline in an entirely different manner than human cells, cleaving its C–N bond to produce trimethylamine (TMA), which is subsequently oxidized to trimethylamine *N*-oxide by hepatic enzymes. This microbial-human co-metabolic pathway has been linked to several diseases, including non-alcoholic fatty liver disease and atherosclerosis. Even though anaerobic microbial choline conversion into TMA has been known for over a century, its genetic and biochemical bases had not been identified prior to our work. This thesis presents the discovery and validation of a gene cluster responsible for anaerobic choline utilization (*cut* gene cluster), as well as the characterization and inhibition of the key TMA-forming enzyme, choline TMA-lyase (CutC).

Chapter 2 details the bioinformatic approach used to discover the *cut* gene cluster, and the genetic, spectroscopic and cultivation-based strategies to connect these genes to anaerobic choline metabolism. We also reveal the involvement of a C–N bond cleaving glycyl radical enzyme (CutC) in

TMA production, an unprecedented reactivity for this enzyme family. Overall, our experimental and computational findings suggest that the *cut* pathway may be a major mechanism for the direct conversion of choline to TMA by gut bacteria.

Chapter 3 describes the in vitro reconstitution and study of choline-TMA lyase CutC and its activating protein, the radical *S*-adenosylmethionine dependent enzyme CutD. We demonstrate that CutC can be activated to the glycyl radical form by CutD under anaerobic conditions, and can process choline to TMA and acetaldehyde with high specificity. Homology modeling and mutagenesis experiments further allow us to conclude that CutC is a glycyl radical enzyme of unique function and a molecular marker for anaerobic choline metabolism.

Chapter 4 presents the results of a close collaboration with the Drennan group at Massachusetts Institute of Technology. They elucidated four high-resolution X-ray structures of wild-type CutC and mechanistically informative mutants in the presence of choline, while I characterized the impact of mutations on choline binding and catalysis. Our data uncover unexpected interactions between the trimethylammonium group of choline and polar amino acids side chains and provide new insight into the mechanism of C–N cleavage by CutC. This work broadens our understanding of radical-based enzyme catalysis and will aid in the rational design of inhibitors of bacterial trimethylamine production.

Chapter 5 depicts initial efforts towards discovering small molecule inhibitors of CutC-mediated choline cleavage. A first round of structure-guided rational design revealed betaine aldehyde as a promising lead. This molecule inhibited choline conversion to TMA by a panel of *cut* gene cluster-containing bacteria and by a human fecal sample. I also describe our high throughput screening approach to identify new inhibitors of choline metabolism by gut microbes, as an orthogonal strategy to rational design. Towards this end, we optimized a media formulation containing choline as sole carbon source, such that survival of gut bacteria grown in this media would be dependent on the *cut* pathway. Overall, our preliminary results show that small molecules can interfere with choline metabolism by anaerobic gut microbes and set the stage for more extensive efforts to discover potent inhibitors of this pathway.

Table of Contents

Acknowledgements

Firstly, I would like to thank my thesis advisor, Prof. Emily P. Balskus, for her guidance and support throughout my doctoral studies, as well as for teaching me everything from biochemical and microbiological techniques to how to think about research. I am also grateful to her for discovering the *cut* gene cluster and for her numerous ideas regarding ways to characterize and inhibit choline metabolism by anaerobic bacteria. In addition, I am thankful to my graduate advising committee members, Prof. Catherine L. Drennan and Prof. Daniel Kahne, for helpful discussions and their support during the past four years. I would also like to thank my undergraduate research mentors, Prof. Kelling Donald and Prof. Wade Downey, for all of their guidance, advice, encouragement, and for introducing me to the world of scientific research.

The work I present in this thesis is the result of several fruitful collaborations with researchers from and outside of the Balskus group. I am very grateful to Dr. Ana Martínez-del Campo, a postdoctoral fellow in the Balskus group, who has been leading most microbiological experiments for my project since 2013. Besides having a major contribution to the work presented in sections 2.2.5 through 2.2.8 of this thesis, Ana has been a great friend, patiently answering all of my questions about bioinformatics, culturing of anaerobic bacteria, or working with fecal samples. I am also thankful to all graduate and undergraduate students I have had the privilege of working with: Jonathan Marks for constructing the homology model for CutC and conducting in silico docking experiments, Alix Chan and Kevin Bu for help with cloning and purifying CutC mutants and truncated variants, Marina Orman and Rebecca Taylor for assistance with LC-MS experiments and bacterial growth assays towards finding inhibitors of choline metabolism, Carina Chittim for help with enrichment culturing, and Benjamin Levin for helpful discussions about characterizing glycyl radical enzymes. All crystallography work presented in this thesis was performed by Dr. Michael Funk, a former graduate student in the laboratory of Prof. Catherine Drennan at MIT. I am extremely grateful to Michael and Prof. Drennan for initiating this collaboration with us, for helping us interpret the results described in Chapter 4, and for sharing with us their expertise in radical enzymology. The Balskus group's long-standing collaboration with the Small Molecule Mass

Spectrometry Facility at Harvard has also benefitted my project tremendously through helpful advice and work performed by Kelly Chatman, Jennifer Wang and Dr. Sunia Trauger.

Since at the beginning of my graduate studies I had no knowledge of the instruments needed to answer my research questions, it is only through the patience and guidance of the following scientists that I managed to gather the data presented in this thesis. I am thus indebted to: Dr. Tejia Zhang and Prof. Alan Saghatelian for teaching me how to use an LC-MS instrument and think about my experiments, Dr. Hitomi Nakamura for helping me with organic synthesis techniques, HPLC and troublesome LC-MS instruments, Matthew Wilding, William Collins, Dr. Shaw Huang, Dr. Megan Brophy, Dr. Yifeng Wei, Dr. Graham Sazama, Prof. Alison Fout, and Prof. Theodore Betley for advice concerning EPR, Prof. Turnbaugh and Dr. Henry Haiser for help and advice regarding anaerobic culturing of bacteria, enrichment culturing, and degenerate PCR, and Dr. Corneliu Bodea for assistance with Matlab calculations and computational work. My work was also enabled by researchers willing to provide us with bacterial strains that harbor the *cut* gene cluster. I acknowledge Joao Carita (Universidade Nova de Lisboa, Lisbon) for providing *D. desulfuricans* ATCC 27774, Lee Krumholz and Peter Bradstock (University of Oklahoma, Norman, OK) for providing *D. alaskensis* G20 wild-type and C10(pB6) strains, Karine Gibbs (Harvard University) for providing *P. mirabilis* strains, and the Stanhope group (Cornell University) for providing *S. dysgalactiae*.

Being one of the first members of the Balskus group was a true honor for me, and despite all the challenges, I will forever cherish all the great memories I have from setting up the lab and helping new members. I am grateful to all present and former Balskus group members for being a part of my family during these five years, for helping me out when things were rough, and for celebrating with me when things were great. I am extremely appreciative of Hitomi Nakamura's friendship and support throughout my graduate studies. I cannot thank her enough for being my sidekick in everything from LC-MS troubleshooting to Zumba classes. I am also grateful to Ben Levin for "having my back", especially regarding maintenance of anaerobic chambers, and to Spencer Peck and Ana Martínez-del Campo for editing my thesis and remaining my friends afterwards.

Finally, I would like to thank my family, the pillars on which I stand: my husband Corneliu Bodea, my mother Doina Craciun, my father Mircea-Sorin Craciun, and my mother-in-law Constanta Bodea. Words cannot describe my gratitude for my husband's love, support, encouragements, and understanding; none of my successes would have been possible without him as "my other half". I am also thankful to my mother-in-law for her words of wisdom and advice that helped me overcome the many challenges of graduate school and job searching. My parents have inspired my love of science and whole-heartedly supported my decision to study in the US even though it meant I would be on a different continent and we would seldom see each other. I am extremely grateful for the sacrifices they have made to enable my academic journey and for their continuous belief in my ability to carve my own path in life.

List of Abbreviations

^{o}C	degree Celsius
DTT	dithiothreitol
EC50	half-maximal effective concentration
equiv	equivalent
g	gram
GC-MS/MS	gas chromatography tandem mass spectrometry
h	hour
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HPLC	high-performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
J	coupling constant
L	liter
LB	Luria-Bertani broth
LC-MS/MS	liquid chromatography tandem mass spectrometry
m	mole
min	minute
NADH	β-nicrotinamide adenine dinucleotide phosphate reduced form
Ni NTA	nickel nitriloacetate
o/n	overnight
OD_{600}	optical density at 600 nm
PCR	polymerase chain reaction
psi	pounds per square inch
rpm	revolutions per minute
s	second
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
UV-Vis	ultra-violet visible

Chapter 1: Introduction to the host-gut microbiota metabolic dialogue and glycyl radical enzymes

1.1: General introduction to the human gut microbiota

Humans exist in symbiosis with a large number of commensal microorganisms (primarily members of the Bacterial kingdom, as well as some Archaea, Eukarya, and their viruses), collectively known as the human microbiota. Of the many communities that comprise our microbiota, the one residing in our gastrointestinal tract has been the subject of numerous studies due to its profound impact on our physiology and health (1). The beneficial functions of our gut microbes include essential vitamin and nutrient synthesis, immune-cell development and homeostasis, digestion of inaccessible food components, fat metabolism support, epithelial homeostasis, enteric nerve regulation, and angiogenesis promotion (2). Disruption of the gut microbiota (dysbiosis) stemming from either environmental factors (*e.g,* antibiotic use, diet, stres, and injury) or mammalian metabolism (*e.g,* immune signaling, hormone synthesis), has been associated with a number of diseases, such as allergy, inflammatory bowel disease (IBD), obesity, nonalcoholic fatty liver disease (NAFLD), colon cancer, and diabetes (3). In addition, the gut microbiota can metabolize therapeutic drugs, altering their safety and efficacy (4), and plays a crucial role in protecting against the establishment of pathogenic microbes (5).

It is currently estimated that the number of bacteria in our bodies is on the same order as the number of human cells, but with 10^{11} bacterial cells/mL, our colon represents an anaerobic bioreactor that is considered the most densely populated microbial habitat on Earth (6). A metagenomic sequencing study of the gut microbiota found that the total number of genes possessed by this community outnumbers human genes ~150 fold (7). This implies that our gut microbes have metabolic and biosynthetic capabilities that greatly surpass those of human cells. Despite numerous reports linking changes in the composition of this microbial community to various health and disease states (8), our current knowledge of specific gut bacterial enzymes, metabolites and pathways remains limited, thus providing exciting opportunities for discovery.

Analyses of the identities of our gut microbes have revealed that this complex community consists of more than 500 species belonging to eight bacterial divisions, with Firmicutes and Bacteroidetes as the dominant phyla (90 % of the distal gut microbiota) (8, 9). The diversity within this community is considered the result of coevolution with the host and a strong selection for specific bacteria whose collective metabolic capacity is beneficial to the host (10). The gut microbiota of an infant is seeded at birth, and it is influenced by a series of factors, including method of delivery, breast feeding and weaning (11). During the early years, the microbiota changes and develops rapidly, reaching a steady state in adulthood. A final set of changes occurs during old age, when the immune system starts to decline (3). Thus, throughout our lives, the function and composition of our gut microbiota are continually influenced not just by external environmental factors, but also by our own genome and metabolism.

Through a substantial high throughput sequencing effort, the Human Microbiome Project (HMP) has provided a comprehensive picture of the human microbiota from multiple body sites, including the gut. Interestingly, researchers from the HMP were unable to identify a set of common gut microbial species shared among all individuals, comparable to our shared genome. Instead, they uncovered that the diversity and abundance of gut microbes highly vary among individuals (12). Nonetheless, metagenomic sequencing performed by the HMP and its European equivalent, the Metagenomics of the Human Intestinal Tract (MetaHIT) Consortium (7), revealed a functional core shared by each individual in the cohort and encoded across many microbial species. This finding implies that the metabolic pathways necessary for human-associated microbial life are conserved, despite the variability in bacterial taxonomic composition. One of the most striking results of these metagenomic studies is the abundance of unknown biochemical functions encoded by the human gut microbiota. Of the assembled HMP stool metagenomes, ~70% of the genes could not be assigned a GO (Gene Ontology) or EC (Enzyme Commission) annotation (13). Thus, a timely characterization of novel biochemical functions from our microbiota is necessary in order to establish causality relationships between microbial metabolites, genes, metabolic pathways and human health.

1.2: Tools to study the gut microbiota

One of the most common techniques used to investigate human-associated microbial communities is the study of cultured isolates. When used in conjunction with whole genome sequencing, this approach enables researchers to link phenotypes to specific microbial genes (4). However, given that a large proportion of gut bacteria remain unculturable, this strategy cannot be applied universally. In order to simulate some of the complexity of the gut environment in vitro, a number of models were developed, ranging from simple single vessel batch cultures to complex multistage pH-regulated continuous cultures (*e.g,* SHIME or TIM-1) (14). These in vitro systems provide a facile and cost-effective means of studying the gut microbiome in one or more gut compartments, but they still lack an epithelial mucosa, as well as immunological and neuroendocrine interactions with a host (14). Recent advances in microfluidics have allowed the development of a human gut-on-a-chip microdevice that was used to coculture multiple commensal microbes in contact with living human intestinal epithelial cells and analyze the effect of peristalsis-associated mechanical deformations on inflammatory bowel disease (15). Despite representing a proof-of-principle device, the human gut-on-a-chip appears as a robust platform for the modeling of human intestinal disease in a modular manner.

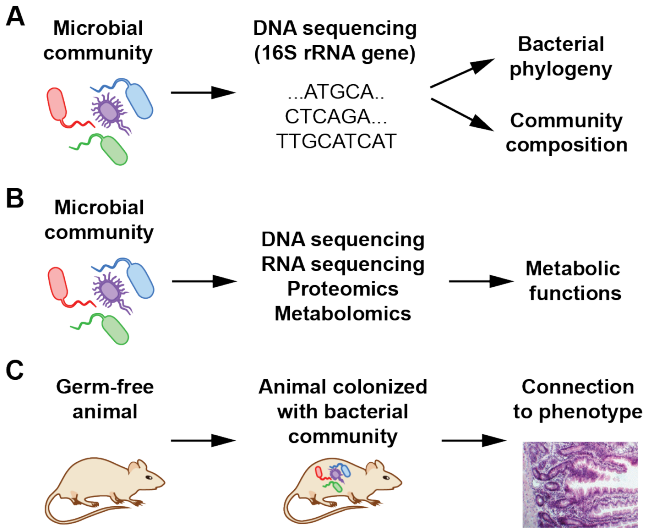


Figure 1.1: Approaches for studying the human gut microbiota. **(A)** Taxonomic profiling provides information about the identities and abundances of different organisms in communities. **(B)** Functional profiling can be performed by determining the total gene content of microbial communities, as well as the abundance of gene transcripts, proteins, and metabolites. **(C)** Gnotobiotic (germ-free) animal models represent a method to connect specific organisms and functions to host phenotypes.

First employed in environmental microbiology, taxonomic profiling (or 16S rRNA sequencing) represents the most widely used method for characterizing the identities and abundances of different bacterial taxa from the gut microbiota, because it uses the variable region of the bacterial 16S rRNA gene as a phylogenetic marker (16) (**Figure 1.1.A**). Optimized for high throughput sequencing by the HMP and MetaHIT, this approach has provided an unprecedented insight into the bacterial diversity of healthy individuals (7, 12), as well as correlations between gut community composition and various health and disease states (2). A substantial limitation of taxonomic profiling is that closely related bacteria often possess different metabolic pathways due to mechanisms such as gene loss and horizontal gene transfer (17), thus limiting our ability to correlate community composition to biochemical function. In addition, taxonomic profiling cannot distinguish between active and inactive, or resident and transient members of the gut microbial community.

In recent years, research concerning host-gut microbiota interactions has focused on the functional profiling of this microbial community through approaches such as metagenomics, metatranscriptomics, metaproteomics, and metabolomics (**Figure 1.1.B**). These methods provide information regarding the total microbial gene abundance, which genes are expressed and active at the moment of sampling, and which metabolites are enriched or depleted in different tissues and bodily fluids (18). Correlating this information with different host and disease states has revealed unexpected potential contributions of the gut microbiota to various host processes (*e.g.,* inflammation, DNA damage, energy homeostasis) (3). However, it is still difficult to determine specific molecular mechanisms though which gut microbes elicit an effect on host health. This obstacle ultimately limits our ability to harness beneficial activities of our gut bacteria or intervene therapeutically in associated diseases.

Another strategy to investigate the impact of gut microbial metabolism on the mammalian host is the development of animal models (mice, zebrafish, etc.) with non-existent or simplified microbiotas, *e.g.,* germ-free (gnotobiotic) mice, antibiotic-treated rodents with a temporary knockout of select bacterial groups, and animals with transplanted microbiota from human patients (**Figure 1.1.C**). By allowing

disruptions of the gut microbiota to be studied within a controlled experimental setting, animal models represent the best method to assess causality and establish mechanistic hypotheses regarding the connection between gut bacterial function and disease (19). Mouse models have proved particularly effective because gnotobiotic mice can support the growth of the majority of human-associated bacterial genera, and one can leverage the comprehensive knowledge of mouse genetics and availability of genetically modified mouse strains (19). However, translating the results of mouse model studies to humans is complicated by intrinsic differences in intestinal anatomy and function, as well as by specific chemical-mediated crosstalk between the host and gut microbes (19).

The large fraction of uncharacterized genes in the gut microbiome drastically hampers our ability to understand and prevent detrimental metabolic functions of this microbial community. Joice et al. (13) proposed to combine the techniques presented above in a pipeline for the identification and validation of unknown biochemical functions of our gut microbiota. Starting with microbial genomic, metagenomic or transcriptomic data, standard sequence analysis methods (homology-based assignment or domain profiling) can be used to assign a putative activity to a gene of unknown function. Additional bioinformatic approaches can further help to refine this prediction. Useful tools include: comparative metagenomics (association of uncharacterized microbial products with characterized genes across samples through the use of data integration), supervised curation (manual establishment of a consensus among multiple complementary automated annotations), phylogenetic profiling (analysis of co-occurrence of genes across isolates), and network context (analysis of isolate co-expression, interaction or functional linkage). Following putative classification of the unknown gene, bioactivity must be validated and characterized by experimental methods, such as microbial culture, heterologous expression of genes, and direct isolation of products. Functional assays (e.g. enzymatic assays, microbial co-culture to determine microbial inhibitory effects, host cell profiling, or assessments of in vivo host phenotype) can then be performed in order to investigate the gene's putative activity. Even though this pipeline is taking advantage of stat-of-the-art approaches for functional assignment and validation, the chemical diversity associated with the gut microbiota is still poorly understood. More than 50% of the genes identified in the

assembled HMP stool metagenomes lack homology to characterized proteins (13), and novel functions are continuously uncovered even for enzymes that can be annotated (20, 21). As a result, there is an urgent need for biochemical studies that characterize novel enzymology of gut bacteria.

1.3: Host-gut microbiota metabolic interactions

The metabolic communication between the host and its gut microbiota is established via several ways: gut microbes biosynthesize unique molecules, process dietary components and pharmaceuticals in a different manner than the host, and engage in co-metabolic pathways (**Figure 1.2**). Possessing expanded chemical capabilities that have evolved to benefit the host, gut bacteria produce valuable nutrients such as vitamins B12, K, biotin, folate, thiamine, riboflavin, and pyridoxine (22). Symbiosis is also facilitated by microbial cellular components such as zwitterionic bacterial capsular polysaccharides (ZPS), which can modulate the host immune system (23). Through their unique metabolism of dietary components, gut bacteria provide the host with the ability to harvest energy from indigestible foods, such as soluble dietary fibers (fructans, pectin, inulin and xylans) and resistant starch. This metabolism leads to the production of short chain fatty acids (SCFAs) such as butyrate, propionate and acetate whose total concentrations in the lumen can reach 80–130 mM under physiological conditions (24). SCFAs serve as a major energy source (~70% of energy intake) for intestinal epithelial cells, and are implicated in various processes beneficial to the host, such as maintaining epithelial barrier function, regulating inflammation through G-coupled protein receptors, diminishing oxidative DNA damage and regulating epithelial cell proliferation (25). Recognition of this impact of gut microbial metabolism on the immune system opened up an entirely new approach to the understanding and treatment of human inflammatory disease. Ongoing research in this area currently focuses on isolation of other immunomodulatory molecules produced by gut microbes, and assessment of the beneficial effects of different probiotics and prebiotics on immune homeostasis (25).

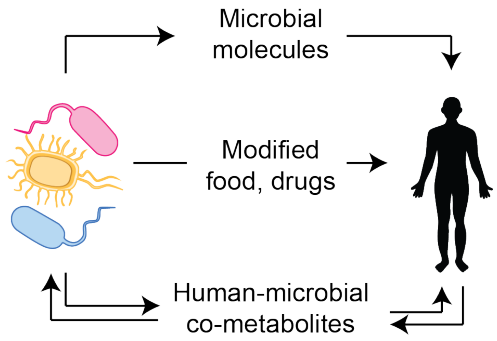


Figure 1.2: Host-gut microbiota metabolic interactions.

Gut bacteria also affect our capacity to metabolize therapeutic drugs. Since the human gut lumen is an anaerobic environment (8), the bacteria that inhabit it are capable of reactions opposite in nature to those performed by human cells (oxidative and conjugative). As a result, gut microbes were found to transform clinical drugs by reduction of double bonds, hydrolysis, dehydroxylation, deacetylation, and deconjugation of glucuronides and sulfates (26). These transformations can have a variety of consequences: drug activation (*e.g.,* sulfasalazine, used for IBD treatment), inactivation (*e.g.,* digoxin, a cardiac glycoside), or severe drug toxicity (*e.g.,* the anticancer drug CPT-11) (26). The detrimental effects of gut microbial drug metabolism on health provided the impetus for the first selective inhibition of a class of nonessential protein targets from gut bacteria, the β-glucuronidases (27). These enzymes remove the glucuronide group from the active form of CPT-11, releasing the DNA-intercalating drug into the gut lumen, which results in severe diarrhea (27). Reducing the gastrointestinal damage caused by in situ gut bacterial reactivation of therapeutic metabolites was also successful in a study in which a microbial β-glucuronidase inhibitor was co-administered with nonsteroidal anti-inflammatory drugs (28). Given the profound impact of gut bacterial metabolism on human health, a molecular understanding of host-gut microbe interactions holds the promise of revealing new therapeutic targets among bacterial proteins. Gut microbial degradation of choline represents an auspicious opportunity due to its association with disease.

1.4: Metabolism of choline by mammalian cells

Choline is an essential nutrient for humans, participating in important processes that extend from fetal development into adulthood and old age (**Figure 1.2**). The most concentrated sources of dietary

choline include liver, eggs, and wheat germ. In foods, choline is found in free or esterified forms, such as phosphocholine, glycerophosphocholine, phosphatidylcholine (PC), and sphingomyelin (29). PC is the major dietary source of choline via its hydrolysis by phospholipase enzymes (type A, C or D) (30), and direct ingestion of PC was shown to increase the level of choline and its metabolites in mice and in humans (31, 32). In the liver, the major fates of PC are secretion into bile and formation of lipoproteins that circulate in the plasma. Approximately 95% of biliary PC is reabsorbed by the intestine, thus exposing the gut microbiota to concentrations of up to 10-20 g of PC/day (33). Besides diet, up to ~15% of choline in the human bloodstream can also be acquired via de novo biosynthesis through the methylation of phosphatidylethanolamine to PC (34) (**Figure 1.3**). However, de novo synthesis of choline alone is insufficient to meet optimal intake requirements, which have been set by the Institute of Medicine at 425 mg/day for women and 550 mg/day for men (29). Once ingested, dietary choline is absorbed by the small intestine, where low affinity choline transporters mediate its uptake (35, 36). Recent studies indicate that choline is recycled in the liver and redistributed from kidney, lung, and intestine to the liver and brain when choline supply is low (29).

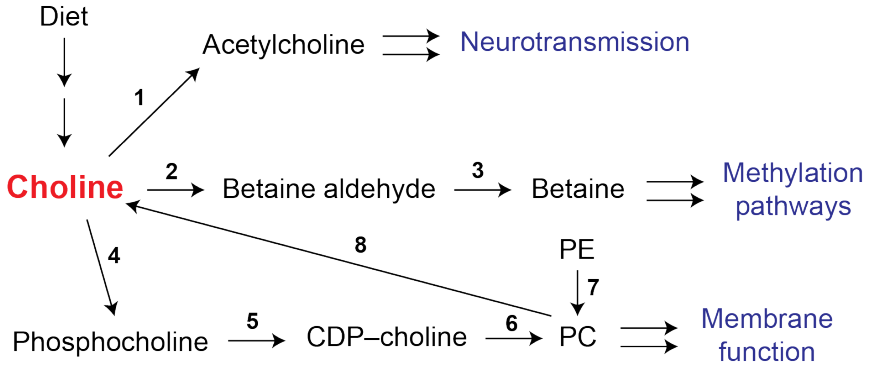


Figure 1.3: Metabolic pathways involving choline. CTP, cytidine 5'-triphosphate; CDP, cytidine 5'-diphosphate; PE, phosphatidylethanolamine; PC, phosphatidylcholine. Enzyme names are indicated by numbers: 1, choline acetyltransferase; 2, choline oxidase; 3, betaine aldehyde dehydrogenase; 4, choline kinase; 5, CTP:phosphocholine cytidylyltransferase; 6, CDP-choline:1,2-diacylglycerol choline phospho-transferase; 7, phosphatidylethanolamine *N*-methyltransferase; 8, phospholipases (type A, C or D).

The primary role of choline is in maintaining cell membrane function by its conversion via the Kennedy pathway in all nucleated cells to PC, the predominant phospholipid (>50%) in most mammalian membranes (29) (**Figure 1.3**). Choline can easily cross the blood-brain barrier and is converted in the

brain to the neurotransmitter acetylcholine, which is involved in sleep, learning, memory and muscle control (37). Choline is also a major dietary source of methyl groups, as it is irreversibly oxidized in the mitochondria of liver and kidney cells to glycine betaine, which participates in the biosynthesis of methionine and *S*-adenosyl methionine (SAM). Because of its wide-ranging role in human metabolism, choline is implicated in a number of diseases including non-alcoholic fatty liver disease, defects in fetal brain development, and cardiovascular disease (via lipoprotein secretion) (29).

1.5: Bacterial metabolism of choline

Due to its abundance in eukaryotic cells, choline is present in many different habitats. Consequently, choline can be metabolized aerobically by a diverse array of environmental bacteria (coryneform bacteria, Cyanobacteria) and opportunistic pathogens (Proteobacteria including *Pseudomonas aeruginosa*) (38, 39). While phosphatidylcholine is ubiquitous in eukaryotic membranes, it is estimated that only 15% of bacteria (primarily Actinomycetales, α- and γ-Proteobacteria) can biosynthesize PC (40). The relative amount of PC detected in different bacterial species can vary widely, ranging from a few percent of the total membrane lipid (0-4% in *Pseudomonas aeruginosa*) to up to 73% in *Acetobacter aceti* (41). Choline is thus catabolized by bacteria under aerobic conditions for the main purpose of generating glycine betaine (GB), a molecule that can enhance bacterial survival in high-salinity environments due to its osmoprotectant properties (42). In the case of pathogens such as *Pseudomonas aeruginosa*, aerobic choline utilization as a carbon and nitrogen source directly regulates virulence factor production. Both glycine betaine and its downstream metabolite, dimethylglycine, induce the transcription of hemolytic phospholipase C (*plcH*), which was found to be important for virulence in several animal models of infection (43, 44).

The most complex bacterial uptake system for choline characterized to date is present in *P. aeruginosa* and involves four transporters that differ in their contributions to growth and osmoprotection in various salt concentrations. CbcXVW and BetT1 contribute most under hypo- and iso-osmolar conditions, whereas BetT2 and BetT3 appear to be the primary transporters under hyperosmolar conditions (45). The metabolic intermediates of aerobic choline catabolism are conserved among the

different bacterial species, and their biosynthesis requires O_2 to proceed. Interestingly, members of at least two different enzyme families can carry out each chemical conversion (**Figure 1.4.A**). The oxidation of choline to form GB can be catalyzed either by monomeric choline oxidases (flavin-dependent) or by the concerted action of an oxidase and an aldehyde dehydrogenase (pyrroloquinoline quinone-dependent) through the intermediacy of betaine aldehyde (46). The demethylation of GB to dimethylglycine can proceed through a zinc-dependent betaine-homocysteine methyltransferase (in Cyanobacteria) or a Rieske-type dioxygenase (in Proteobacteria) (47). Further demethylation to sarcosine and glycine is catalyzed by flavin-dependent oxidases that exist either as single-subunit (Arthrobacter, Bacillus), or multi-subunit (Pseudomonads) enzymes (48, 49). Due to the contribution of choline catabolism to bacterial pathogenesis in organisms such *P. aeruginosa*, current efforts are aimed at developing small molecule inhibitors that would be useful tools for studying this pathway in clinical and environmental isolates and potential new therapeutics (50).

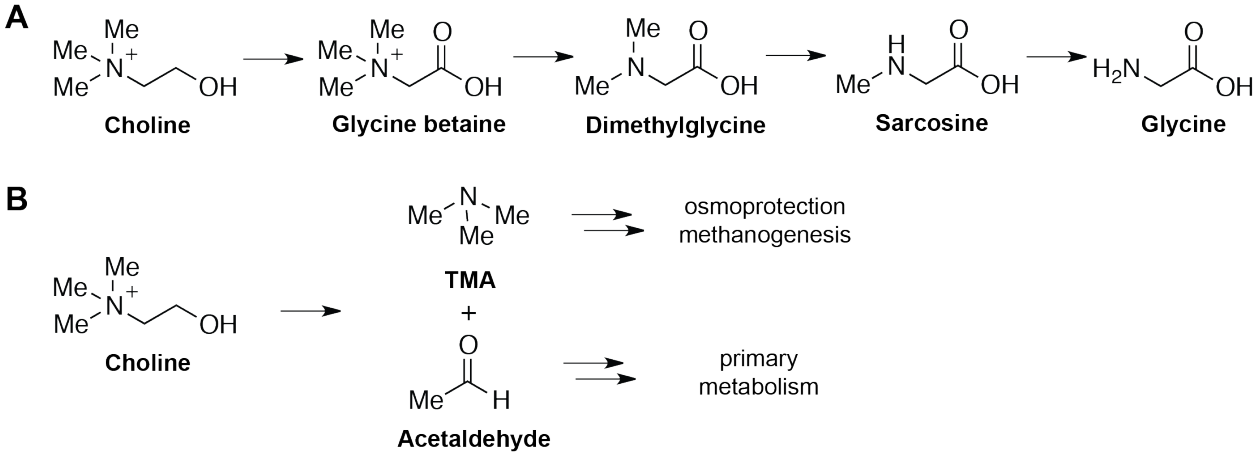


Figure 1.4: Choline catabolic pathways in **(A)** aerobic and **(B)** anaerobic bacteria.

Under strictly anaerobic conditions, certain environmental (51, 52), human-associated (53), and vertebrate-associated (54) bacteria process choline in a different manner, by converting it to trimethylamine (TMA) and acetaldehyde. This metabolic pathway was first reported in 1910 (55), and subsequently found to exist in both facultative and obligate anaerobes, such as *Desulfovibrio desulfuricans*, *Aerobacter aerogenes*, *Shigella alkalescens*, *Proteus mirabilis*, and *Streptococcus sanguis* (56-59). By generating acetaldehyde, anaerobic choline catabolism allows bacteria to utilize choline as a

carbon source (56), while TMA can be used as a substrate for methanogenesis by Archaea within both the gastrointestinal tract of ruminants (51) and marine sediments (52). In the human body, most of the TMA produced in the gut is transported via the portal circulation to the liver, where flavin-dependent monooxygenase enzymes (FMOs) (particularly FMO3), which oxidize it to the water-soluble metabolite trimethylamine *N*-oxide (TMAO) (**Figure 1.5**). Upon entering circulation, TMAO is predominantly excreted by the kidneys into urine (60). Although TMA has been considered a normal constituent of human urine since the beginning of the past century, the microbial origin of the metabolite was not conclusively established until 1992 (54).

Figure 1.5: Choline metabolism to TMAO within the human body.

The pathway for anaerobic bacterial choline metabolism was first studied by Hayward and Stadtman in 1959 (56). Using mud from North Carolina, they isolated a species of *Vibrio* (later reclassified as *Desulfovibrio desulfuricans*) that was able to ferment choline anaerobically and produce TMA. Their subsequent reports revealed that the first step in choline catabolism is a C–N bond cleavage reaction, and that the resulting co-product acetaldehyde is converted into ethanol and acetate for utilization as a carbon source (61). Using bacterial lysates able to convert choline into TMA (from *Desulfovibrio*, *Clostridia*, *Proteus* and *Streptococcus*), several studies (56-59) attempted to identify the enzymes responsible for C–N cleavage via protein purification, but were unsuccessful. However, these efforts reached some common conclusions. For instance, conversion of choline to TMA was found to require both particulate and soluble protein fractions, potassium ions and a bound divalent cation, such as Fe^{2+}. This activity was improved by conducting enzymatic assays under a nitrogen atmosphere and by the addition of sulfhydryl-reducing agents, but was lost upon all attempts at protein purification. In addition, the reaction was consistently found to be independent of the presence of vitamin B_{12}. This result was quite unexpected because ethanolamine ammonia-lyase, a microbial enzyme that catalyzes a chemically similar

reaction to choline cleavage (ethanolamine conversion into ammonia and acetaldehyde), was found to utilize vitamin B_{12} as cofactor (62). Thus, the genetic and biochemical basis of anaerobic choline metabolism remained a mystery prior to our work.

1.6: Health consequences of gut microbial metabolism of choline

Multiple human diseases are associated with gut microbial TMA production. Studies in humans revealed that the most significant dietary precursors of this microbial metabolite are choline, L-carnitine (primarily from red meat), and TMAO (primarily from fish) (31, 63, 64). The metabolic disorder trimethylaminuria is caused by functional defects in FMO3, the liver enzyme responsible for the detoxification and elimination of TMA through oxidation to the odorless TMAO (60). This genetic condition is inherited in an autosomal recessive manner as a consequence of mutations in the FMO3 gene (65). While the accumulation of TMA is not known to have a deleterious effect on physical health, it often causes social and psychological debilitation due to TMA's distinctive odor of decaying fish being present in urine, sweat or breath, saliva, and other bodily fluids (66). Since many people suffering from trimethylaminuria remain undiagnosed for long periods of time, the incidence of this disorder is not precisely known. However, genetic studies have estimated the incidence of heterozygous carriers to range from 0.5% to 11% depending on the ethnicity examined, with overrepresentation among females of African descent (67, 68). Approaches for managing TMA accumulation include restricting the intake of choline-containing foods and marine fish, ingesting activated charcoal that can absorb TMA, intermittent use of antibiotics to reduce gut bacterial load, and personal hygienic measures (*e.g,* low pH soaps, lotions) (67). However, these treatment options are not universally efficacious for trimethylaminuria.

Low-choline diets are associated with health problems in humans and in mouse models, including NAFLD, neural tube defects, hepatic cancer, and an increased risk of breast cancer (29). Since NAFLD affects ~ 30% of the US population and the mechanisms underlying this condition are incompletely understood (69), several studies in mice and humans have investigated the contribution of gut microbial metabolism to choline deficiency as it relates to this disorder (70, 71). Using ^{1}H NMR spectroscopy-based plasma and urine metabolomics in a mouse model susceptible to fatty liver disease and insulin resistance,

Dumas et al. described a gut microbiota-mediated mechanism underlying the development of fatty liver. Feeding a high-fat diet to mice that are genetically predisposed to NAFLD resulted in low plasma phosphatidylcholine concentrations, mimicking choline deficiency (70). The authors concluded that host genetics could lead to a gut microbiota abundant in choline-utilizing bacteria that ultimately reduce choline bioavailability. More recently, Le Roy et al. demonstrated that NAFLD is transmissible to germ-free mice via microbiota transplant (72). In a different report, Spencer et al. present the results of a 2-month inpatient study of 15 female subjects who were placed on well-controlled diets in which choline levels were manipulated. Taxonomic profiling was used to characterize gut microbiota composition in stool samples (71). Even though each individual's gut microbial community remained distinct over the course of the study, restricted availability of dietary choline was correlated with increased levels of Gammaproteobacteria and Erysipelotrichi, as well as increases in the amount of liver fat (71). By identifying bacterial biomarkers of fatty liver disease resulting from choline deficiency, this study further contributed to the accumulating evidence that gastrointestinal microbes have a role in metabolic disorders.

An additional, unexpected connection between TMA and human health is the recent association of high levels of TMAO in plasma and urine with susceptibility to both chronic kidney disease and cardiovascular disease (31, 73). In a study of over 500 patients with chronic kidney disease and 3,000 controls, TMAO levels were found to predict future mortality risk after adjustment for traditional risk factors (73). In addition, dietary choline (1% wt/wt of chow) and TMAO (0.12% wt/wt of chow) exposure appeared to contribute to renal injury and dysfunction in conventionally housed mice (with intact gut microbiota) (73). These results point towards a direct involvement of TMAO in kidney disease development and progression, but the molecular details of this involvement of are still unknown.

The correlation between TMAO levels in plasma and urine and risk for cardiovascular disease was discovered in an untargeted metabolomics study of almost 2,000 American subjects undergoing elective cardiac evaluation, and it was further supported by a study of over 4,000 subjects undergoing elective coronary angiography (31, 74). TMAO levels were found to predict the risk of major adverse cardiovascular events independently of traditional cardiovascular risk factors, renal function, and

medication use (74). Hazen and co-workers conducted several studies in which deuterated PC (believed to be the main source of choline in our diet (32)) and L-carnitine were fed to both mice and humans, and the gut microbiota was suppressed through the use of antibiotics. Their results confirmed that gut microbes play an obligatory role in generating TMAO from these trimethylammonium-containing nutrients, and revealed that TMAO can directly contribute to aortic plaque formation in mice (31, 63, 74). However, these conclusions remain controversial, as recent studies of over 10,000 Japanese men and women (75), and over 300 Austrian patients (76) failed to observe correlations between either choline intake or TMAO levels and cardiovascular disease mortality risk. It is likely that complex interactions between diet, host genetics and gut microbes have lead to these discrepancies. Consequently, it is of great importance to determine the molecular mechanisms by which TMAO influences all these host processes.

Known to function as osmolyte in marine fish and as small molecule chaperone for protein folding, TMAO was believed until recently to be devoid of biological activity in humans (77, 78). No molecular targets are known for TMAO within the human body and the conclusions drawn from recent experiments with mice disagree regarding the involvement of this small molecule in atherosclerotic plaque development (31, 79). Plaque lesions are typically initiated by foam cell formation, which is characterized by the engorgement of arterial macrophages by oxidized low-density lipoprotein (LDL) cholesterol (80). One study that supplemented the diet of female apolipoprotein E–null ($ApoE^{-/-}$) (atherosclerosis prone) mice with a single high dose of TMAO (0.12% wt/wt of chow) observed enhanced macrophage foam cell formation in both the artery wall and the peritoneal cavity in 3 out of 11 mice (31). While investigating the reasons behind this effect, researchers observed that protein surface levels of previously implicated scavenger receptors (CD36 and SR-A1) were augmented on macrophages involved in cholesterol accumulation and foam cell formation (31). TMAO also elicited significant reductions in reverse cholesterol transport and altered the bile acid pool size and composition in this mouse model (63). A follow-up study in mice and human cell lines (primary aortic endothelial cells and vascular smooth muscle cells) showed that TMAO can promote recruitment of activated leukocytes to endothelial cells and induce inflammation by activating the mitogen-activated protein kinase (MAPK) and nuclear factor kB

(NF-kB) signaling cascade (81). Recent findings also from Hazen and co-workers have further strengthened the connection between TMAO and cardiovascular disease by determining that TMAO has a direct contribution to thrombosis by enhancing platelet aggregation in mice fed TMAO- or choline-supplemented diets (82). In addition, treating washed platelets isolated from mice with TMAO was found to alter inositol-1,4,5-triphosophate (IP3) signaling, and result in Ca^{2+} release from intracellular stores, thus increasing platelet responsiveness (82). Despite this novel insight into the connection between TMAO and cardiovascular disease, the molecular targets or receptors for TMAO still remain unknown.

A different study with apolipoprotein E–null male mice transfected with human cholesteryl ester transfer protein (hCETP) determined that TMAO and L-carnitine are likely to provide a protective effect on atherosclerosis development, and do not impact plasma lipid and lipoprotein levels (79). This mouse strain could be a better model for atherosclerosis because CETP is a key enzyme in reverse cholesterol transport, transferring cholesterol ester from high-density lipoprotein (HDL) to LDL cholesterol (83). The use of male mice also avoided the variability and high expression levels of FMO isoforms reported in female mice. Additionally, this study examined the dose-dependent effect of TMAO on mouse macrophages, and found that TMAO does not contribute to foam cell formation, even at concentrations 10-fold higher than the reported human plasma maximum concentration (79). The differences in diet of this mouse model (atherogenic, with L-carnitine amounts in the range seen in dietary supplements) compared to the diet of mice used by Hazen and co-workers (normal chow supplemented with 5-fold higher levels of L-carnitine) (63) might account for the discrepancy in their results. Nonetheless, follow-up studies are needed to verify the accuracy of the claims made by either group of researchers.

A recent investigation of TMAO-promoted atherosclerosis in multiple strains of inbred mice revealed that the levels of this metabolite explain 11% of the variation in the disease, and that FMO3 expression is induced by dietary bile acids via a mechanism involving the farnesoid X receptor (FXR), a bile acid-activated nuclear receptor (84). A mechanistic link between FMO3, cholesterol, and bile acid metabolism recently emerged during a transcriptional profiling study in mouse models of altered cholesterol transport. Knockdown of FMO3 was found to reorganize whole-body cholesterol balance and

exacerbate downstream inflammatory responses in the liver (85). Another study of FMO3 knockdown in insulin-resistant mice concluded that FMO3 is required for the expression of a central node for metabolic control, transcription factor FoxO1, and that it is increased in obese/insulin-resistant humans, thus confirming the central role of this enzyme in metabolic dysfunction (86). Despite the controversy regarding the direct effect of TMAO on aortic plaque development, these studies have opened the possibility that FMO3 could have regulatory functions distinct from its enzymatic activity. It also seems likely that TMAO levels might influence FMO3 regulation, which highlights the importance of studying TMAO formation within the human body.

1.7: Glycyl radical enzymes play key roles in anaerobic microbial metabolism

The glycyl radical enzyme (GRE) family has emerged as an important class of catalysts for anaerobic microorganisms from the environment as well as from the human body. These biocatalysts perform chemically difficult reactions through the utilization of a glycine-centered radical generated by a separate activating enzyme (AE or activase), a [4Fe-4S] cluster containing member of the radical S-adenosylmethionine (SAM or AdoMet) protein family (87). GREs play central roles in anoxic microbial metabolic pathways, such as nucleotide reduction (class III ribonucleotide reductase (RNR)), carbon-source utilization (pyruvate formate-lyase (PFL), benzylsuccinate synthase (BSS), glycerol dehydratase (GDH)) or biosynthesis of bacteriostatic compounds (4-hydroxyphenylacetate decarboxylase (HPD)) (**Figure 1.6**). The impact of GRE-catalyzed reactions on human health has only recently begun to be appreciated. Several metagenomic studies have found GREs to be among the most abundant protein families in the human gut microbiota and strongly overrepresented in this environment relative to other microbial habitats (88, 89). These observations are also supported by proteomic data revealing GREs to be highly abundant in proteomes from human intestinal samples (90).

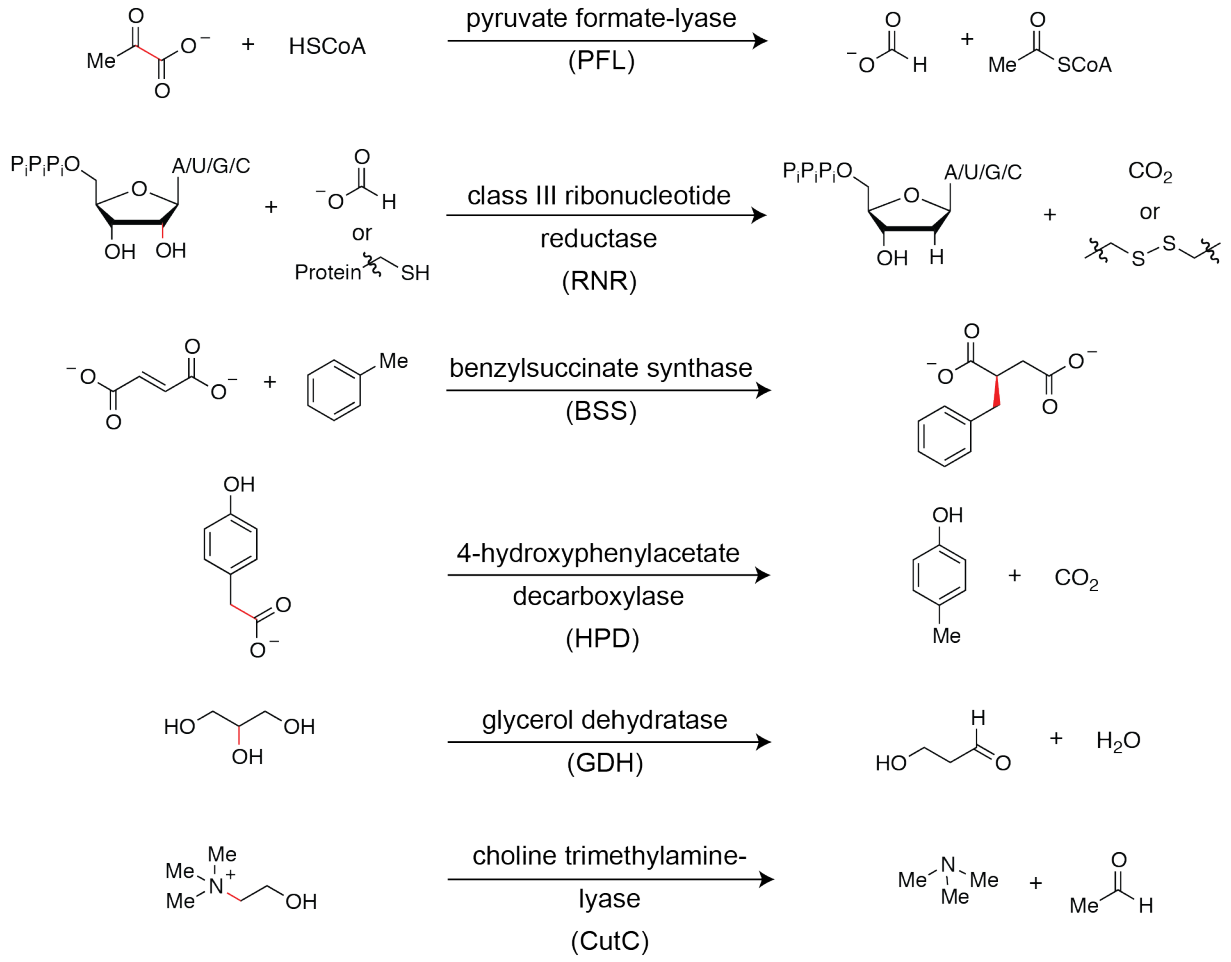


Figure 1.6: Functionally characterized GRE classes. P_i, phosphate; A, adenine; U, uracil; G, guanine, C, cytosine.

As of 2016, six classes of GREs have been functionally verified and biochemically characterized (**Figure 1.6**):

Pyruvate formate-lyase (PFL, EC 2.3.1.54) was the first enzyme found to contain a glycyl radical via electron paramagnetic resonance spectroscopy (EPR) (91, 92) by the Knappe group at Heidelberg University. PFL is one of the key enzymes in prokaryotic glycolysis. It catalyzes the homolytic C–C bond cleavage of pyruvate to produce formate and acetyl-CoA, two central metabolites that can be used in biosynthesis or to generate ATP, the energy currency of the cell (93). One of the most studied members of the GRE superfamily, PFL from *E. coli* was the first GRE found to be overexpressed and post-translationally modified by the activating enzyme (PFL-AE) only under anaerobic conditions

(94). Although the *E. coli* genome encodes another α-ketoacid formate lyase that can catalyze the same reaction (TdcE), knockout of the PFL gene resulted in severe anaerobic growth impairment (95). In line with these results, the PFL-catalyzed pyruvate cleavage by the non-recombinant enzyme is very efficient, with k_{cat} values of 1100 s^{-1} for the forward reaction, and 260 s^{-1} for the reverse reaction, with the equilibrium favoring the forward step (K_{eq} ~ 750) (96).

Class III or anaerobic ribonucleotide reductase (class III RNR, EC 1.17.4.2) catalyzes the anaerobic reduction of ribonucleotide triphosphates (NTPs) to 2'-deoxynucleotides (dNTPs), the building blocks of DNA, with the oxidation of either formate (97) or an active site thiol pair (98). Ribonucleotide reduction is the only known pathway for de novo synthesis of deoxyribonucleotides and it is required in all living organisms. This reaction is accomplished by three different classes of ribonucleotide reductases hypothesized to have evolved from the same common ancestor and diversified depending on the presence of oxygen in the surrounding environment (99). Even though they rely on different cofactors, all RNRs generate an active site thiyl radical that initiates the reduction chemistry. The class I RNRs use cofactors resulting from the reaction of reduced metals (Fe, Mn, and Fe/Mn) with O_2 and are present only in aerobic organisms, whereas the class II RNRs use adenosylcobalamin in an O_2-independent reaction and are present in both aerobes and anaerobes (100). The class III RNRs are distributed throughout facultative and obligate anaerobic bacteria and are also found in some archaea, primarily strictly anaerobic methanogens (98). The activity of an O_2-sensitive RNR was first detected in extracts of the strictly anaerobic methanogenic archaeon, *Methanobacterium marburgensis* (101) and later found in *E. coli* (102). Similarities between PFL and class III RNR were noted soon after these discoveries, and investigations of the mechanism of glycyl radical generation and activity proceeded in parallel for these two classes of GREs (102). The most well characterized class III RNRs to date are from *Escherichia coli*, *Lactococcus lactis*, *Thermotoga maritima*, and T4 bacteriophage.

X-succinate synthases (XSSs) are a class of GREs responsible for functionalizing a variety of hydrocarbons (toluene, 2-methylnaphthalene, cresols, n-alkanes) via addition of a terminal or subterminal aliphatic C–H bond across the double bond of fumarate, initiating the decontamination of hydrocarbon-

polluted environments (103). The resulting succinate derivatives further undergo thioesterification, carbon-skeleton rearrangement (or ring cleavage), decarboxylation and beta-oxidation, resulting in the production of either acetate or propionate, depending on whether the produced fatty acids are even or odd numbered (104). The most well studied member of this class is **benzylsuccinate synthase (BSS, EC 4.1.99.11)** from *Thauera aromatica*, which catalyzes the addition of toluene to fumarate, generating benzylsuccinate (105, 106). BSS is one of a few GREs composed of multiple subunits, as it was isolated as a stable $(\alpha\beta\gamma)_2$ heterohexamer. Biochemical and structural studied revealed that the α-subunit contains the glycyl radical and is responsible for the catalytic activity, and binding of toluene and fumarate, whereas the accessory subunits β and γ display folds related to high potential iron–sulfur proteins, and appear to play redox and structural roles (105). BSSγ binds distally to the active site and buries a hydrophobic region of BSSα, while BSSβ is proposed to modulate the conformational dynamics required for activity by binding to a hydrophilic surface of BSSα proximal to the active site (105). Despite the successful purification and crystallization of the BSS complex, BSS-AE has resisted purification attempts and the activity has not yet been reconstituted and studied in vitro. XSS activity has been observed in a variety of anaerobic phototrophic bacteria, as well as in metal-ion-reducing *Geobacter metalireducens* and in syntrophic communities of methanogenic archaea and sulfate-reducing bacteria (107-109), thus underscoring the importance of these metabolic pathways for bacterial survival in hydrocarbon-rich environments.

4-Hydroxyphenylacetate decarboxylase (HPD, EC 4.1.1.83) is primarily present in commensal and pathogenic Clostridia and catalyzes the last step in the biosynthesis of the bacteriostatic metabolite *p*-cresol (4-methylphenol), decarboxylation of 4-hydroxyphenylacetate, which is derived from tyrosine (110-112). *P*-cresol production by *Clostridium difficile* might contribute to virulence, since these bacteria can withstand 35-fold greater concentrations of *p*-cresol than other strains (113). Structural studies of HPD revealed that, similarly to BSS, HPD is composed of a main catalytic subunit (β) harboring the

glycyl radical, and a small subunit (γ) that binds two [4Fe-4S] clusters and is believed to be involved in regulating the oligomeric state and activity of the heterooctameric complex $(\beta\gamma)_4$.

Glycerol dehydratase (GDH, EC 4.2.1.30) catalyzes the first step in glycerol utilization as a carbon source by *Clostridium butyricum*, dehydrating glycerol to 3-hydroxypropionaldehyde, which can be further reduced by aldehyde dehydrogenases from the organism to the valuable polyester precursor 1,3-propanediol (114, 115). Recent structural and biochemical efforts have primarily focused on characterizing GDH from *Clostridium butyricum* (115, 116), but other diol dehydrating GREs have been uncovered. A transcriptional profiling study found that a GRE homologous to GDH mediates the dehydration of propanediol during growth of human-associated gut bacteria such as *Roseburia inulinivorans* on host-derived fucose (117). This metabolic pathway eventually leads to the production of propionate, a short chain fatty acid that is beneficial for the host (1). A second GRE-containing gene cluster also upregulated in the presence of L-fucose was found in the environmental organism *Clostridium phytofermentans* (118), but the distribution of GREs from this class has not been yet investigated.

Choline TMA-lyase (CutC, EC 4.3.99.4) catalyzes choline cleavage to trimethylamine and acetaldehyde in bacteria from the human gut and the environment. Its discovery, characterization, distribution, and inhibition will be described in this thesis.

1.8: Common structural characteristics of glycyl radical enzymes

It is believed that the selection of a simple glycine residue as a radical storage element represents an ancient strategy in the evolution of life (99, 119). The conversion of RNA- to DNA-based life was likely performed by a common ancestor of the three classes of ribonucleotide reductase, most closely related to class III RNR (120). Bioinformatic analyses suggest that this class is the most primitive RNR, while the other types of RNRs arose by duplication followed by divergence, as they adapted to increasing levels of oxygen in the Earth's atmosphere. (120). Since PFL is a key enzyme in anaerobic metabolism of glucose, it likely appeared early in the evolution of life, perhaps even predating RNR (119). Interestingly, formate, the product of the PFL reaction, is the reductant used by many members of the class III RNR.

The similarities in chemistry and structural features between these two enzymes despite lower sequence homology (99) thus suggest that PFL and class III RNRs are evolutionarily related, but have undergone divergent evolution.

Crystallographic and biochemical studies of GREs such as PFL, BSS, HPD, and class III RNR have revealed that GREs conserved structural features that facilitate catalysis (87). Their common barrel architecture has been named the "GRE fold": a ten-stranded β/α barrel with two loops containing catalytic residues inserted into the center of the barrel (**Figure 1.7.A**). It is hypothesized that this structural motif shields radicals generated in the active site from solvent, similar to the case of adenosylcobalamin (AdoCbl) enzymes or radical SAM enzymes (121, 122). The glycine residue that is converted into the glycyl radical is close to the C-terminus (the glycyl radical domain) and is part of a loop (Gly loop) comprised of a highly conserved stretch of amino acids (RVXG). A second loop (Cys loop) containing a conserved Cys residue is located opposite to the Gly loop and is deeply buried in the barrel such that the two residues can contact each other and participate in a radical relay (**Figure 1.7.B**). The active site of each GRE is typically located on or in the vicinity of the Cys loop, such that a direct interaction could be established between the Cys residue and the substrate (87).

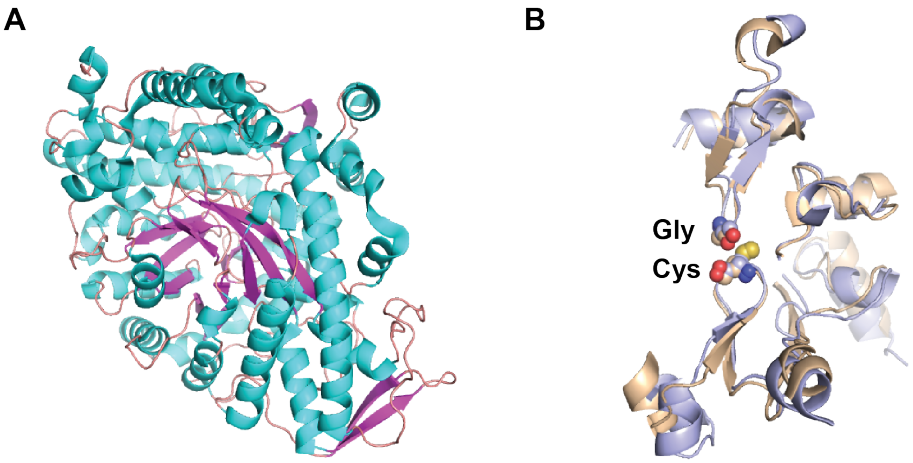


Figure 1.7: (A) General ten-stranded β/α barrel architecture of glycyl radical enzymes (the structure shown here is that of *E. coli* PFL, PDB ID: 2PFL). **(B)** The Gly and Cys loops contain the Gly residue used for radical storage and the Cys residue used for initiating catalysis. Shown is a superimposition of the structures for *E. coli* PFL (PDB ID: 2PFL) and *C. butyricum* glycerol dehydratase (PDB ID: 1R9D).

1.9: Mechanistic features of glycyl radical enzymes

The initial activation step of all GREs involves the installation of the glycyl radical by the activating enzyme (AE) (123). This activation process is thought to be specific for each functional class of GREs, since GREs and AEs are typically encoded adjacently in microbial genomes. To the best of our knowledge, an AE from one class can only activate GREs possessing the same activity as its native partner. All AEs are members of the radical SAM enzyme superfamily, and they utilize a reduced [4Fe-4S]$^{+}$ cluster to convert SAM into the 5'-deoxyadenosyl radical (5'-dA•) (**Figure 1.8**). Due to the extreme oxygen sensitivity of the [4Fe-4S] cluster and the instability of many AEs that have incomplete clusters, it is very challenging to express these enzymes in heterologous hosts and to reconstitute their clusters. The only AE crystal structure available is that of PFL-AE, which is also the most extensively characterized AE (124). Similar to other radical SAM enzymes, the active site of PFL-AE contains three Cys residues in a conserved CX_3CX_2C motif that coordinate a [4Fe-4S] cluster. The Fe atom of the cluster that does not interact directly with the protein is instead coordinated by the amino nitrogen and carboxyl oxygen of SAM in a square-pyramidal geometry that allows the bottom of the pyramid to be occupied by the sulfonium group of SAM. This arrangement allows inner sphere electron transfer between the reduced [4Fe-4S] cluster in the +1 oxidation state to the C–S antibonding σ∗ orbital of SAM (125). The electron transfer lowers the bond dissociation energy (BDE) of the $C_{adenosine}$–S in SAM (~60 kcal/mol prior to electron transfer) such that homolytic bond cleavage can occur to form 5'-dA• and methionine (126). The radical 5'-dA• is extremely reactive (C–H BDE ~99 kcal/mol), and thus capable of abstracting the C–H bond of the GRE active site glycine (C–H BDE ~79.3 kcal/mol) (**Figure 1.8**) (127).

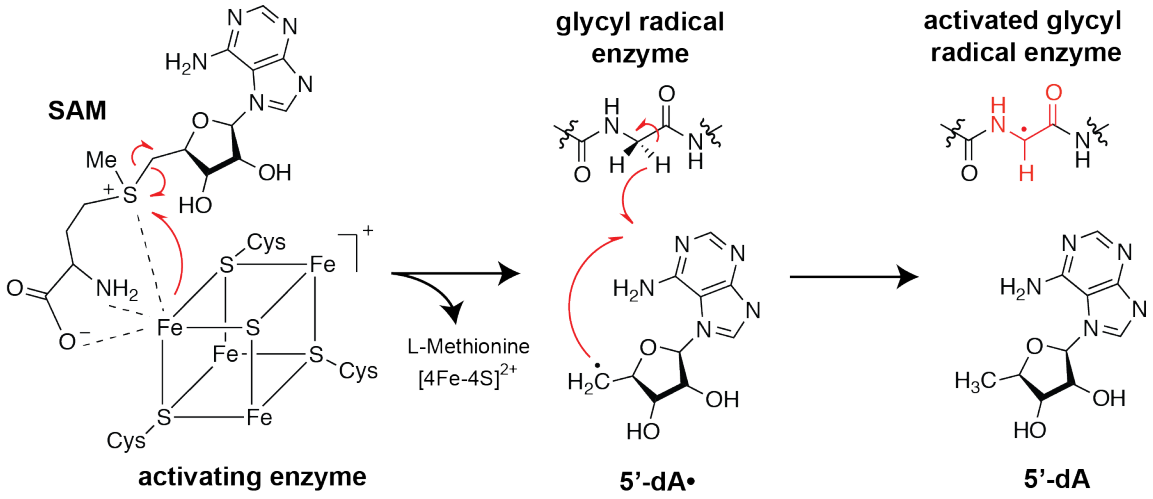


Figure 1.8: General scheme for activation of a glycyl radical enzyme by a GRE activating enzyme. SAM: S-adenosyl methionine, 5'-dA: 5'-deoxyadenosine.

The glycyl radical is stabilized via the captodative effect stemming from its adjacency to the electron donating amide nitrogen and electron withdrawing carbonyl, and is believed to be the "storage form" of the radical (110). Early EPR studies of PFL performed by the Knappe group determined that the radical in activated PFL is present on an sp^2-hybridized α-carbon atom of a specific glycine residue (91, 92). When handled under anaerobic conditions, all GREs display a similar EPR signal, a doublet with hyperfine coupling of 1.4–1.5 mT (due to the interaction of the glycyl radical with the nucleus of the α-hydrogen), centered at g = 2.003–2.004 (92, 111, 128-130). This specific EPR signal allows the quantification of the extent of activation in vitro (thus enabling kinetic measurements), as well as the detection of GRE activity in whole cells overexpressing the enzyme during growth on a specific substrate, as in the case of methylpentylsuccinate synthase (131). Under aerobic conditions, the glycyl radical of PFL was found to react with oxygen to form α-hydroxyglycine, which leads to the irreversible cleavage of the polypeptide backbone (132). This susceptibility of the glycyl radical to irreversible damage by oxygen implies that activated GREs need to be handled under strictly anaerobic conditions, and explains why GREs are not present in obligate aerobic organisms. Interestingly, facultative anaerobic bacteria such as *E. coli* contain a small protein similar in sequence to the glycyl radical domain of PFL (YfiD), which was found to bind oxygen damaged PFL and rescue its activity upon activation by PFL-AE (133). It is

currently not known whether other glycyl radical enzymes also utilize this "spare part" strategy. However, as more classes of GREs are functionally characterized, other strategies of preventing or mitigating oxidative damage are likely to be uncovered.

Even though a crystal structure of a GRE-AE complex has not been yet elucidated, Vey and coworkers were able to crystallize PFL-AE bound to a small peptide containing the Gly loop sequence conserved in all GREs (124). Interestingly, the peptide is bound such that the Gly residue was adjacent (4.1 Å) to SAM (**Figure 1.9**). In this conformation, the 5'-dA• resulting from SAM cleavage would be perfectly poised to abstract the pro-*S* hydrogen of the Gly residue in accordance with previous experiments (134). In all crystal structures of GREs determined to date, this Gly residue is deeply buried 8-20 Å from the surface of the protein (87). This paradox has led to a model of activation for PFL in which the Gly loop undergoes a conformational change such that it can contact the AE. The model was proposed in a study characterizing mixtures of PFL and PFL-AE using circular dichroism (CD) spectroscopy and EPR, which concluded that the presence of the AE induces an "open" conformation of PFL that makes the glycyl radical more susceptible to quenching by external reductants (dithiothreitol) (135). For all the GREs characterized thus far, the maximum extent of activation obtained in vitro is 0.5 moles of glycyl radical/mol of polypeptide (129, 136). Since most GREs exist as dimeric proteins (87), it is hypothesized that activation of one monomer precludes the activation of the other through a conformational change, despite the ~70 Å distance between the two glycyl radical domains. Many questions thus remain concerning GRE activation, particularly regarding the enzyme's strategy for stabilizing the glycyl radical, the stoichiometry of the process, and its regulation.

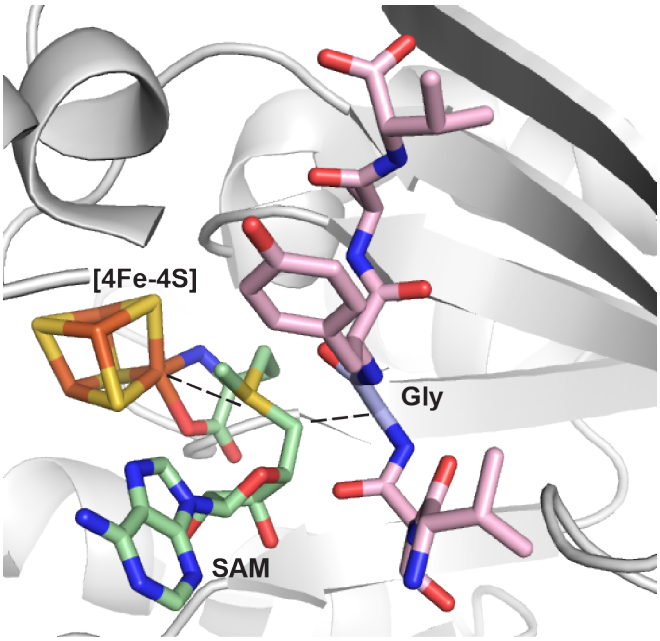


Figure 1.9: Crystal structure of pyruvate formate-lyase (PDB ID: 3CB8) activating enzyme in complex with SAM (green) and a Gly loop peptide (pink and blue) that mimics the glycyl radical domain of PFL.

In all GREs, the glycyl radical is believed to function as a site of radical storage. Structural and mechanistic studies of GREs led to the hypothesis that upon substrate binding, the glycyl radical generates a thiyl radical on the cysteine residue of the Cys loop (87). Even though this thiyl radical has not yet been directly observed by spectroscopic methods, it is believed to be readily accessible, since the S–H BDE was calculated at ~81.7 kcal/mol (137). In addition, the protein might modulate the reactivity of the glycyl radical via conformational control to ensure that thiyl radical formation is not rate-determining (137). The conservation of these catalytic Gly and Cys residues in all known GREs and their close proximity in all elucidated GRE structures support a general mechanism for GRE catalysis: substrate binding to the activated GRE could induce the generation of the thiyl radical, which would initiate the radical reaction and would be regenerated upon product formation (**Figure 1.10**).

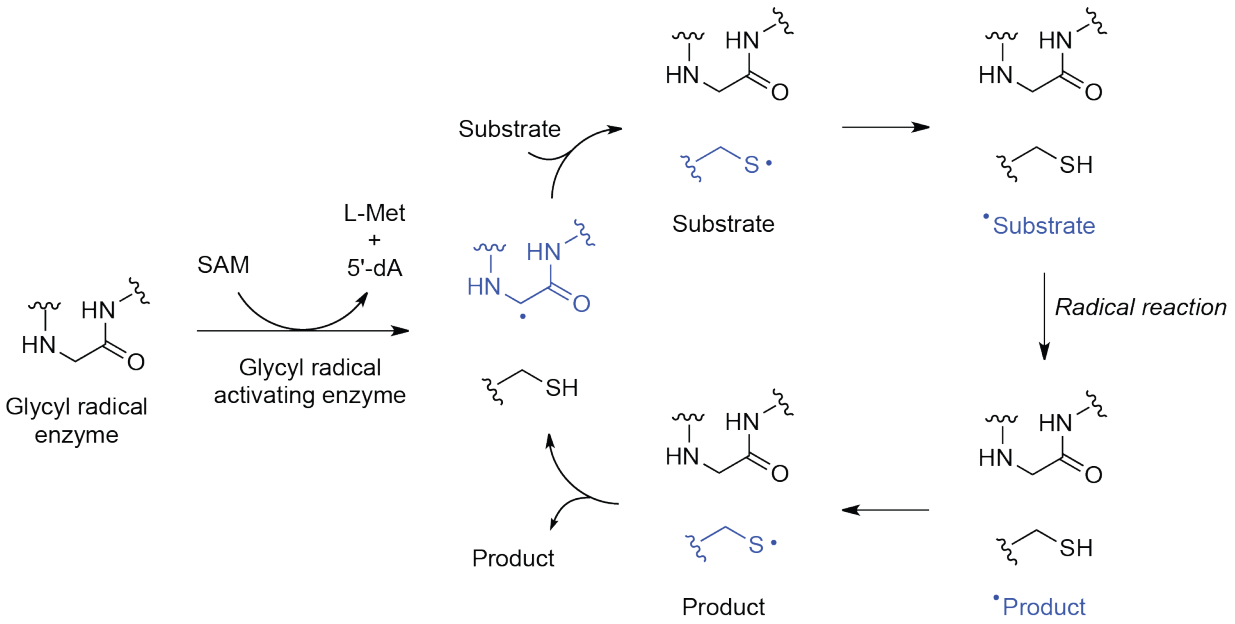


Figure 1.10: Shared mechanistic hypothesis for GRE function.

Mechanistic and spectroscopic studies of GRE-catalyzed reactions have proven very challenging to perform, since these reactions are extremely oxygen sensitive and tend to proceed very quickly. As a result, researchers have proposed mechanisms for the function of each class of GREs based on the crystal structures of unactivated GREs and subsequent computational studies (**Figure 1.11-1.15**). In the case of PFL (**Figure 1.11**), two Cys residues are present in the active site and were found to participate in the reaction (138). Upon conversion of Cys419 (from *E. coli* PFL) into a thiyl radical by hydrogen atom abstraction by the glycyl radical on Gly734, the adjacent Cys residue, Cys418, would be converted into a thiyl radical. This radical species could interact with pyruvate to form a tetrahedral intermediate that would yield a formyl radical and acylated Cys418 upon collapsing. The formyl radical could regenerate the thiyl radical at Cys419, which would convert CoA into a thiyl radical. This species could interact with the acylated Cys418 to form a thiyl radical at this residue and acetyl-CoA. For class III RNR, ribonucleotide reduction would be initiated by formation of a 3'-nucleotide radical by the active site thiyl radical (98) (**Figure 1.12**). Loss of water to form a 3'-ketyl radical could be facilitated by formate (139), and reduction of this species to the 3'-keto-deoxy-nucleotide was found to be accompanied by the formation of a thiosulfuranyl radical between the bottom face cysteine thiyl radical and a methionine

residue (140). The thiosulfuranyl radical is proposed to convert formate into a radical species that could participate in the reduction of the 3'-keto-deoxy-nucleotide via proton-coupled electron transfer using an unknown source of protons (98). Based on the crystal structure of BSS in complex with its substrates, it is believed that the reaction catalyzed by this enzyme would be initiated by H-atom abstraction from toluene to generate a benzyl radical (106) (**Figure 1.13**). This resulting species could then react with fumarate to form a benzylsuccinyl radical intermediate, which would regenerate the thiyl radical to produce (*R*)-benzylsuccinate. A computational study of the HPD mechanism proposed that the reaction would be initiated through radical transfer between the catalytic cysteine and the substrate carboxylic group and a proton transfer from the hydroxyl group of the substrate to a glutamate residue (141) (**Figure 1.14**). Decarboxylation could then be coupled with proton transfer from a glutamate residue to the phenolic hydroxyl group to afford a p-hydroxybenzyl radical, which would regenerate the thiyl radical. In the case of GDH, a computational study proposed that glycerol dehydration could occur directly from a substrate-based radical with the assistance of a general base (Glu) and a general acid (His) present in the active site (**Figure 1.15**). The resulting product radical would then abstract a hydrogen atom from the active site Cys to reform the thiyl radical. It is thus apparent that for each class of GREs, active site residues are perfectly poised to control the chemistry of each radical reaction.

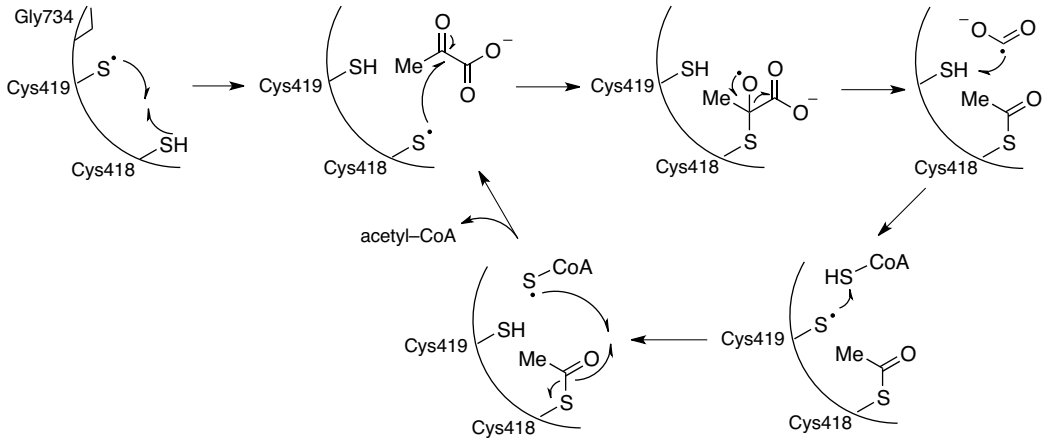


Figure 1.11: Proposed mechanism for the radical reaction of PFL. Amino acids are numbered corresponding to PFL from *E. coli*.

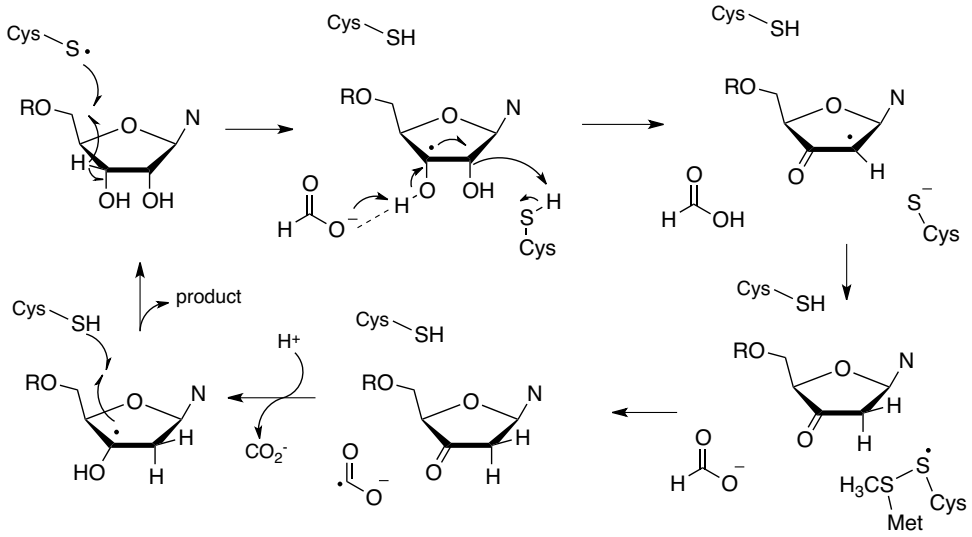


Figure 1.12: Proposed mechanism for the radical reaction of class III RNR. R = di- or triphosphate, N = nucleobase.

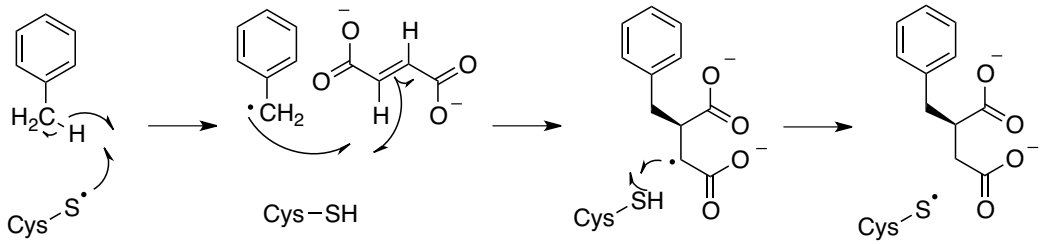


Figure 1.13: Proposed mechanism for the radical reaction of BSS.

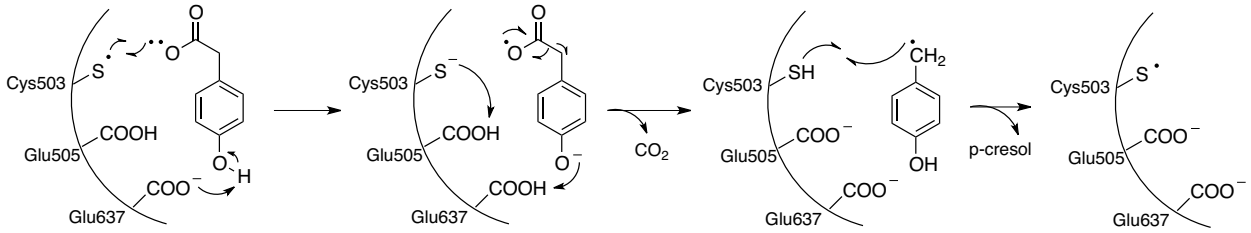


Figure 1.14: Proposed mechanism for the radical reaction of HPD. Amino acids are numbered corresponding to HPD from *Clostridium scatologenes.*

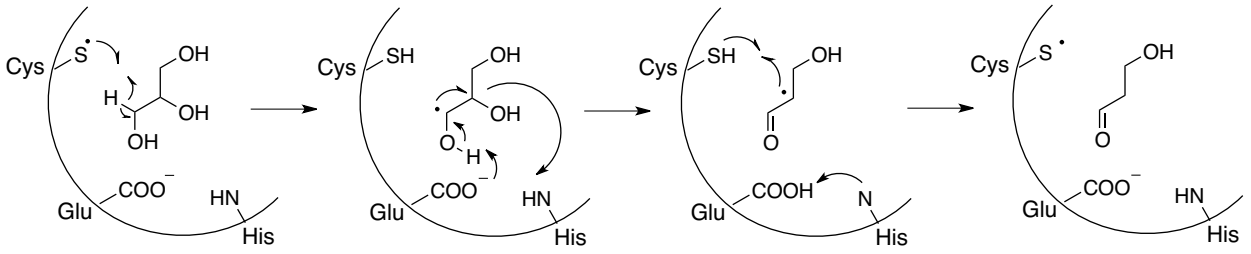


Figure 1.15: Proposed mechanism for the radical reaction of GDH.

Bioinformatic analyses of GREs in sequenced genomes of both obligate and facultative anaerobic bacteria performed by Lehtiö and Goldman (20) and by Balskus group members (manuscript in preparation) have revealed many GREs of unknown function, thus highlighting the functional diversity of this enzyme superfamily. Given the prevalence of GREs in the human gut (88) and their susceptibility to misannotation due to their high amino acid sequence similarity (with the exception of class III RNR) (20), there is a great need to understand the molecular basis for the activities of these enzymes. This knowledge will enable predictions of the types of metabolism performed by uncharacterized GREs, helping to prioritize them for future study. The identification of specific conserved residues involved in substrate binding and catalysis will also facilitate future bioinformatic analyses aimed at characterizing the distribution of specific GREs in microbial genomes and metagenomes.

1.10: References

1. Sharon G*, et al.* (2014) Specialized metabolites from the microbiome in health and disease. *Cell Metab* 20(5):719-730.

2. Holmes E, Li JV, Athanasiou T, Ashrafian H, & Nicholson JK (2011) Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol* 19(7):349-359.

3. Nicholson JK*, et al.* (2012) Host-gut microbiota metabolic interactions. *Science* 336(6086):1262-1267.

4. Haiser HJ*, et al.* (2013) Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* 341(6143):295-298.

5. Buffie CG*, et al.* (2015) Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* 517(7533):205-208.

6. Sender R, Fuchs S, & Milo R (2016) Revised estimates for the number of human and bacteria cells in the body. *bioRxiv*.

7. Qin J*, et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59-65.

8. Donaldson GP, Lee SM, & Mazmanian SK (2016) Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* 14(1):20-32.

9. Eckburg PB*, et al.* (2005) Diversity of the human intestinal microbial flora. *Science* 308(5728):1635-1638.

10. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, & Gordon JI (2005) Host-bacterial mutualism in the human intestine. *Science* 307(5717):1915-1920.

11. Dominguez-Bello MG*, et al.* (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* 107(26):11971-11975.

12. Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207-214.

13. Joice R, Yasuda K, Shafquat A, Morgan XC, & Huttenhower C (2014) Determining microbial products and identifying molecular targets in the human microbiome. *Cell Metab* 20(5):731-741.

14. Williams CF*, et al.* (2015) Comparative analysis of intestinal tract models. *Annu Rev Food Sci Technol* 6:329-350.

15. Kim HJ, Li H, Collins JJ, & Ingber DE (2016) Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. *Proc Natl Acad Sci USA* 113(1):E7-E15.

16. Fox GE, Magrum LJ, Balch WE, Wolfe RS, & Woese CR (1977) Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Natl Acad Sci USA* 74(10):4537-4541.

17. Smillie CS*, et al.* (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480(7376):241-244.

18. van Baarlen P, Kleerebezem M, & Wells JM (2013) Omics approaches to study host-microbiota interactions. *Curr Opin Microbiol* 16(3):270-277.

19. Nguyen TL, Vieira-Silva S, Liston A, & Raes J (2015) How informative is the mouse for human gut microbiota research? *Dis Model Mech* 8(1):1-16.

20. Lehtio L & Goldman A (2004) The pyruvate formate lyase family: sequences, structures and activation. *Protein Eng Des Sel* 17(6):545-552.

21. Haft DH & Basu MK (2011) Biological systems discovery in silico: radical *S*-adenosylmethionine protein families and their target peptides for posttranslational modification. *J Bacteriol* 193(11):2745-2755.

22. LeBlanc JG*, et al.* (2013) Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotechnol* 24(2):160-168.

23. Avci FY & Kasper DL (2010) How bacterial carbohydrates influence the adaptive immune system. *Annu Rev Immunol* 28:107-130.

24. Cummings JH, Pomare EW, Branch WJ, Naylor CPE, & Macfarlane GT (1987) Short chain fatty-acids in human large-intestine, portal, hepatic and venous-blood. *Gut* 28(10):1221-1227.

25. Maslowski KM & Mackay CR (2011) Diet, gut microbiota and immune responses. *Nat Immunol* 12(1):5-9.

26. Sousa T*, et al.* (2008) The gastrointestinal microbiota as a site for the biotransformation of drugs. *International Journal of Pharmaceutics* 363(1-2):1-25.

27. Wallace BD*, et al.* (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* 330(6005):831-835.

28. LoGuidice A, Wallace BD, Bendel L, Redinbo MR, & Boelsterli UA (2012) Pharmacologic targeting of bacterial beta-glucuronidase alleviates nonsteroidal anti-inflammatory drug-induced enteropathy in mice. *J Pharmacol Exp Ther* 341(2):447-454.

29. Zeisel SH & da Costa KA (2009) Choline: an essential nutrient for public health. *Nutr Rev* 67(11):615-623.

30. Lewis ED, Field CJ, & Jacobs RL (2015) Should the forms of dietary choline also be considered when estimating dietary intake and the implications for health? *Lipid Technology* 27(10):227-230.

31. Wang Z*, et al.* (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57-63.

32. Miller CA*, et al.* (2014) Effect of egg ingestion on trimethylamine-N-oxide production in humans: a randomized, controlled, dose-response study. *Am J Clin Nutr* 100(3):778-786.

33. Northfield TC & Hofmann AF (1975) Biliary lipid output during three meals and an overnight fast. 1. Relationship to bile-acid pool size and cholesterol saturation of bile in gallstone and control subjects. *Gut* 16(1):1-12.

34. Sherriff JL, O'Sullivan TA, Properzi C, Oddo JL, & Adams LA (2016) Choline, its potential role in nonalcoholic fatty liver disease, and the case for human and bacterial genes. *Adv Nutr* 7(1):5-13.

35. Tso P & Fujimoto K (1991) The absorption and transport of lipids by the small intestine. *Brain Res Bull* 27(3-4):477-482.

36. Michel V, Yuan ZF, Ramsubir S, & Bakovic M (2006) Choline transport for phospholipid synthesis. *Exp Biol Med* 231(5):490-504.

37. Zeisel SH (2006) The fetal origins of memory: The role of dietary choline in optimal brain development. *J Pediatr* 149(5):S131-S136.

38. Wargo MJ (2013) Homeostasis and catabolism of choline and glycine betaine: lessons from *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 79(7):2112-2120.

39. Kortstee GJ (1970) The aerobic decomposition of choline by microorganisms. I. The ability of aerobic organisms, particularly coryneform bacteria, to utilize choline as the sole carbon and nitrogen source. *Arch Mikrobiol* 71(3):235-244.

40. Aktas M*, et al.* (2010) Phosphatidylcholine biosynthesis and its significance in bacteria interacting with eukaryotic cells. *Eur J Cell Biol* 89(12):888-894.

41. Geiger O, Lopez-Lara IM, & Sohlenkamp C (2013) Phosphatidylcholine biosynthesis and function in bacteria. *Biochim Biophys Acta* 1831(3):503-513.

42. Boch J, Kempf B, & Bremer E (1994) Osmoregulation in *Bacillus subtilis*: synthesis of the osmoprotectant glycine betaine from exogenously provided choline. *J Bacteriol* 176(17):5364-5371.

43. Rahme LG*, et al.* (1995) Common virulence factors for bacterial pathogenicity in plants and animals. *Science* 268(5219):1899-1902.

44. Jander G, Rahme LG, & Ausubel FM (2000) Positive correlation between virulence of *Pseudomonas aeruginosa* mutants in mice and insects. *J Bacteriol* 182(13):3843-3845.

45. Malek AA, Chen CL, Wargo MJ, Beattie GA, & Hogan DA (2011) Roles of three transporters, CbcXWV, BetT1, and BetT3, in *Pseudomonas aeruginosa* choline uptake for catabolism. *J Bacteriol* 193(12):3033-3041.

46. Landfald B & Strom AR (1986) Choline-glycine betaine pathway confers a high-level of osmotic tolerance in *Escherichia coli*. *J Bacteriol* 165(3):849-855.

47. Meskys R, Harris RJ, Casaite V, Basran J, & Scrutton NS (2001) Organization of the genes involved in dimethylglycine and sarcosine degradation in *Arthrobacter spp.*: implications for glycine betaine catabolism. *Eur J Biochem* 268(12):3390-3398.

48. Wargo MJ, Szwergold BS, & Hogan DA (2008) Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism. *J Bacteriol* 190(8):2690-2699.

49. Wagner MA, Khanna P, & Jorns MS (1999) Structure of the flavocoenzyme of two homologous amine oxidases: monomeric sarcosine oxidase and N-methyltryptophan oxidase. *Biochemistry* 38(17):5588-5595.

50. Fitzsimmons LF*, et al.* (2011) Small-molecule inhibition of choline catabolism in *Pseudomonas aeruginosa* and other aerobic choline-catabolizing bacteria. *Appl Environ Microbiol* 77(13):4383-4389.

51. Neill AR, Grime DW, & Dawson RMC (1978) Conversion of choline methyl groups through trimethylamine into methane in the rumen. *J Biochem* 170(3):529-535.

52. King GM (1984) Metabolism of trimethylamine, choline, and glycine betaine by sulfate-reducing and methanogenic bacteria in marine sediments. *Appl Environ Microb* 48(4):719-725.

53. Bain MA, Fornasini G, & Evans AM (2005) Trimethylamine: metabolic, pharmacokinetic and safety aspects. *Curr Drug Metab* 6(3):227-240.

54. al-Waiz M, Mikov M, Mitchell SC, & Smith RL (1992) The exogenous origin of trimethylamine in the mouse. *Metabolism* 41(2):135-136.

55. Ackermann D & Schutze H (1910) The formation of trimethylamine by *Bacterium prodigiosum*. *Zentralb Physiol* 24:210-211.

56. Hayward HR & Stadtman TC (1959) Anaerobic degradation of choline. I. Fermentation of choline by an anaerobic, cytochrome-producing bacterium, *Vibrio cholinicus* n. sp. *J Bacteriol* 78:557-561.

57. Bradbeer C (1965) Clostridial fermentations of choline and ethanolamine. I. Preparation and properties of cell-free extracts. *Journal of Biological Chemistry* 240(12):4669-&.

58. Sandhu SS & Chase T (1986) Aerobic degradation of choline by *Proteus mirabilis*: enzymatic requirements and pathway. *Can J Microbiol* 32(9):743-750.

59. Chao CK & Zeisel SH (1990) Formation of trimethylamine from dietary choline by *Streptococcus sanguis* I, which colonizes the mouth. *J Nutr Biochem* 1(2):89-97.

60. Krueger SK & Williams DE (2005) Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol Therapeut* 106(3):357-387.

61. Hayward HR & Stadtman TC (1960) Anaerobic degradation of choline. II. Preparation and properties of cell-free extracts of *Vibrio cholinicus*. *J Biol Chem* 235:538–543.

62. Kaplan BH & Stadtman ER (1968) Ethanolamine deaminase, a cobamide coenzyme-dependent enzyme. II. Physical and chemical properties and interaction with cobamides and ethanolamine. *J Biol Chem* 243(8):1794-&.

63. Koeth RA*, et al.* (2013) Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* 19(5):576-585.

64. Zhang AQ, Mitchell SC, & Smith RL (1999) Dietary precursors of trimethylamine in man: A pilot study. *Food Chem Toxicol* 37(5):515-520.

65. Dolphin CT, Janmohamed A, Smith RL, Shephard EA, & Phillips IR (1997) Missense mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour syndrome. *Nat Genet* 17(4):491-494.

66. Christodoulou J (2012) Trimethylaminuria: An under-recognised and socially debilitating metabolic disorder. *J Paediatr Child Health* 48(3):E153-E155.

67. Messenger J, Clark S, Massick S, & Bechtel M (2013) A review of trimethylaminuria: (fish odor syndrome). *J Clin Aesthet Dermatol* 6(11):45-48.

68. Mitchell SC, Zhang AQ, Barrett T, Ayesh R, & Smith RL (1997) Studies on the discontinuous N-oxidation of trimethylamine among Jordanian, Ecuadorian and New Guinean populations. *Pharmacogenetics* 7(1):45-50.

69. Noga AA, Zhao Y, & Vance DE (2002) An unexpected requirement for phosphatidylethanolamine N-methyltransferase in the secretion of very low density lipoproteins. *Journal of Biological Chemistry* 277(44):42358-42365.

70. Dumas ME*, et al.* (2006) Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci U S A* 103(33):12511-12516.

71. Spencer MD*, et al.* (2011) Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology* 140(3):976-986.

72. Le Roy T*, et al.* (2013) Intestinal microbiota determines development of non-alcoholic fatty liver disease in mice. *Gut* 62(12):1787-1794.

73. Tang WHW*, et al.* (2015) Gut microbiota-dependent trimethylamine N-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circ Res* 116(3):448-455.

74. Tang WH*, et al.* (2013) Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 368(17):1575-1584.

75. Nagata C*, et al.* (2015) Choline and betaine intakes are not associated with cardiovascular disease mortality risk in Japanese men and women. *J Nutr* 145(8):1787-1792.

76. Mueller DM*, et al.* (2015) Plasma levels of trimethylamine-N-oxide are confounded by impaired kidney function and poor metabolic control. *Atherosclerosis* 243(2):638-644.

77. Bennion BJ & Daggett V (2004) Counteraction of urea-induced protein denaturation by trimethylamine N-oxide: a chemical chaperone at atomic resolution. *Proc Natl Acad Sci USA* 101(17):6433-6438.

78. Raymond JA (1994) Seasonal variations of trimethylamine oxide and urea in the blood of a cold-adapted marine teleost, the rainbow smelt. *Fish Physiol Biochem* 13(1):13-22.

79. Collins HL*, et al.* (2016) L-Carnitine intake and high trimethylamine N-oxide plasma levels correlate with low aortic lesions in ApoE(-/-) transgenic mice expressing CETP. *Atherosclerosis* 244:29-37.

80. Yu XH, Fu YC, Zhang DW, Yin K, & Tang CK (2013) Foam cells in atherosclerosis. *Clin Chim Acta* 424:245-252.

81. Seldin MM, *et al.* (2016) Trimethylamine N-oxide promotes vascular inflammation through signaling of mitogen-activated protein kinase and nuclear factor-κB. *J Am Heart Assoc* 5(2).

82. Zhu W, *et al.* (2016) Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* 165(1):111-124.

83. Barter PJ, *et al.* (2003) Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis. *Arterioscler Thromb Vasc Biol* 23(2):160-167.

84. Bennett BJ, *et al.* (2013) Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell Metab* 17(1):49-60.

85. Warrier M, *et al.* (2015) The TMAO-generating enzyme flavin monooxygenase 3 is a central regulator of cholesterol balance. *Cell Rep* (10):326-338.

86. Miao J, *et al.* (2015) Flavin-containing monooxygenase 3 as a potential player in diabetes-associated atherosclerosis. *Nat Commun* 6:6498.

87. Selmer T, Pierik AJ, & Heider J (2005) New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem* 386(10):981-988.

88. Ellrott K, Jaroszewski L, Li WZ, Wooley JC, & Godzik A (2010) Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput Biol* 6(6):e1000798.

89. Kurokawa K, *et al.* (2007) Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* 14(4):169-181.

90. Kolmeder CA, *et al.* (2012) Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLoS One* 7(1):e29913.

91. Unkrig V, Neugebauer FA, & Knappe J (1989) The free radical of pyruvate formate-lyase. Characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analog hypophosphite. *Eur J Biochem* 184(3):723-728.

92. Wagner AFV, Frey M, Neugebauer FA, Schafer W, & Knappe J (1992) The free-radical in pyruvate formate-lyase is located on glycine-734. *Proc Natl Acad Sci USA* 89(3):996-1000.

93. Knappe J & Sawers G (1990) A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. *Fems Microbiol Rev* 75(4):383-398.

94. Sawers G & Bock A (1988) Anaerobic regulation of pyruvate formate-lyase from *Escherichia coli* K-12. *J Bacteriol* 170(11):5330-5336.

95. Sawers G, Hesslinger C, Muller N, & Kaiser M (1998) The glycyl radical enzyme TdcE can replace pyruvate formate-lyase in glucose fermentation. *J Bacteriol* 180(14):3509-3516.

96. Knappe J, Blaschkowski HP, Grobner P, & Schmitt T (1974) Pyruvate formate-lyase of *Escherichia coli*: the acetyl-enzyme intermediate. *Eur J Biochem* 50(1):253-263.

97. Mulliez E, Ollagnier S, Fontecave M, Eliasson R, & Reichard P (1995) Formate is the hydrogen donor for the anaerobic ribonucleotide reductase from *Escherichia coli*. *Proc Natl Acad Sci USA* 92(19):8759-8762.

98. Wei YF*, et al.* (2014) The class III ribonucleotide reductase from *Neisseria bacilliformis* can utilize thioredoxin as a reductant. *Proc Natl Acad Sci USA* 111(36):E3756-E3765.

99. Stubbe J, Ge J, & Yee CS (2001) The evolution of ribonucleotide reduction revisited. *Trends Biochem Sci* 26(2):93-99.

100. Stubbe J (1998) Ribonucleotide reductases in the twenty-first century. *Proc Natl Acad Sci USA* 95(6):2723-2724.

101. Sprengel G & Follmann H (1981) Evidence for the Reductive Pathway of Deoxyribonucleotide Synthesis in an Archaebacterium. *FEBS Lett* 132(2):207-209.

102. Reichard P (1993) The anaerobic ribonucleotide reductase from *Escherichia coli*. *J Biol Chem* 268(12):8383-8386.

103. Acosta-Gonzalez A, Rossello-Mora R, & Marques S (2013) Diversity of benzylsuccinate synthase-like (*bssa*) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering. *Appl Environ Microb* 79(12):3667-3676.

104. Callaghan AV (2013) Enzymes involved in the anaerobic oxidation of n-alkanes: from methane to long-chain paraffins. *Front Microbiol* 4(89).

105. Funk MA, Judd ET, Marsh ENG, Elliott SJ, & Drennan CL (2014) Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proc Natl Acad Sci USA* 111(28):10161-10166.

106. Funk MA, Marsh ENG, & Drennan CL (2015) Substrate-bound structures of benzylsuccinate synthase reveal how toluene is activated in anaerobic hydrocarbon degradation. *J Biol Chem* 290(37):22398-22408.

107. Kimes NE*, et al.* (2013) Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front Microbiol* 4(50).

108. Kane SR, Beller HR, Legler TC, & Anderson RT (2002) Biochemical and genetic evidence of benzylsuccinate synthase in toluene-degrading, ferric iron-reducing *Geobacter metallireducens*. *Biodegradation* 13(2):149-154.

109. Jaekel U, Zedelius J, Wilkes H, & Musat F (2015) Anaerobic degradation of cyclohexane by sulfate-reducing bacteria from hydrocarbon-contaminated marine sediments. *Front Microbiol* 6(116).

110. Selmer T & Andrei PI (2001) p-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycyl radical enzyme catalysing the formation of p-cresol. *Eur J Biochem* 268(5):1363-1372.

111. Yu L, Blaser M, Andrei PI, Pierik AJ, & Selmer T (2006) 4-Hydroxyphenylacetate decarboxylases: properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* 45(31):9584-9592.

112. Martins BM*, et al.* (2011) Structural basis for a Kolbe-type decarboxylation catalyzed by a glycyl radical enzyme. *J Am Chem Soc* 133(37):14666-14674.

113. Dawson LF, Stabler RA, & Wren BW (2008) Assessing the role of p-cresol tolerance in *Clostridium difficile*. *J Med Microbiol* 57(6):745-749.

114. Raynaud C, Sarcabal P, Meynial-Salles I, Croux C, & Soucaille P (2003) Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc Natl Acad Sci USA* 100(9):5010-5015.

115. O'Brien JR*, et al.* (2004) Insight into the mechanism of the B_{12}-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* 43(16):4635-4645.

116. Feliks M & Ullmann GM (2012) Glycerol dehydratation by the B_{12}-independent enzyme may not involve the migration of a hydroxyl group: a computational study. *J Phys Chem B* 116(24):7076-7087.

117. Scott KP, Martin JC, Campbell G, Mayer CD, & Flint HJ (2006) Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium *Roseburia inulinivorans*. *J Bacteriol* 188(12):4340-4349.

118. Petit E*, et al.* (2013) Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. *PLoS One* 8(1).

119. Reichard P (1997) The evolution of ribonucleotide reduction. *Trends Biochem Sci* 22(3):81-85.

120. Torrents E, Aloy P, Gibert I, & Rodriguez-Trelles F (2002) Ribonucleotide reductases: divergent evolution of an ancient enzyme. *J Mol Evol* 55(2):138-152.

121. Dowling DP, Vey JL, Croft AK, & Drennan CL (2012) Structural diversity in the AdoMet radical enzyme superfamily. *Biochim Biophys Acta* 1824(11):1178-1195.

122. Dowling DP, Croft AK, & Drennan CL (2012) Radical use of Rossmann and TIM barrel architectures for controlling coenzyme B_{12} chemistry. *Annu Rev Biophys* 41:403-427.

123. Shisler KA & Broderick JB (2014) Glycyl radical activating enzymes: structure, mechanism, and substrate interactions. *Arch Biochem Biophys* 546:64-71.

124. Vey JL*, et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.

125. Dey A*, et al.* (2011) S K-edge XAS and DFT calculations on SAM dependent pyruvate formate-lyase activating enzyme: nature of interaction between the Fe_4S_4 cluster and SAM and its role in reactivity. *J Am Chem Soc* 133(46):18656-18662.

126. Fontecave M, Atta M, & Mulliez E (2004) S-adenosylmethionine: nothing goes to waste. *Trends Biochem Sci* 29(5):243-249.

127. Himo F & Eriksson LA (1998) Catalytic mechanism of pyruvate formate-lyase (PFL). A theoretical study. *J Am Chem Soc* 120(44):11449-11455.

128. Sun XY*, et al.* (1996) The free radical of the anaerobic ribonucleotide reductase from *Escherichia coli* is at glycine 681. *Journal of Biological Chemistry* 271(12):6827-6831.

129. Mulliez E, Padovani D, Atta M, Alcouffe C, & Fontecave M (2001) Activation of class III ribonucleotide reductase by flavodoxin: a protein radical-driven electron transfer to the iron-sulfur center. *Biochemistry* 40(12):3730-3736.

130. Krieger CJ, Roseboom W, Albracht SPJ, & Spormann AM (2001) A stable organic free radical in anaerobic benzylsuccinate synthase of *Azoarcus sp* strain T. *Journal of Biological Chemistry* 276(16):12924-12927.

131. Rabus R*, et al.* (2001) Anaerobic initial reaction of n-alkanes in a denitrifying bacterium: evidence for (1-methylpentyl)succinate as initial product and for involvement of an organic radical in n-hexane metabolism. *J Bacteriol* 183(5):1707-1715.

132. Reddy SG*, et al.* (1998) Dioxygen inactivation of pyruvate formate-lyase: EPR evidence for the formation of protein-based sulfinyl and peroxyl radicals. *Biochemistry* 37(2):558-563.

133. Wagner AFV*, et al.* (2001) YfiD of *Escherichia coli* and Y061 of bacteriophage T4 as autonomous glycyl radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase. *Biochem Biophys Res Commun* 285(2):456-462.

134. Frey M, Rothe M, Wagner AFV, & Knappe J (1994) Adenosylmethionine-dependent synthesis of the glycyl radical in pyruvate formate-lyase by abstraction of the glycine C-2 pro-*S* hydrogen atom - studies of [^{2}H]glycine-substituted enzyme and peptides homologous to the glycine-734 site. *J Biol Chem* 269(17):12432-12437.

135. Peng Y, Veneziano SE, Gillispie GD, & Broderick JB (2010) Pyruvate formate-lyase, evidence for an open conformation favored in the presence of its activating enzyme. *J Biol Chem* 285(35):27224-27231.

136. Crain AV & Broderick JB (2013) Flavodoxin cofactor binding induces structural changes that are required for protein-protein interactions with NADP(+) oxidoreductase and pyruvate formate-lyase activating enzyme. *Biochim Biophys Acta* 1834(12):2512-2519.

137. Himo F & Siegbahn PEM (2003) Quantum chemical studies of radical-containing enzymes. *Chem Rev* 103(6):2421-2456.

138. Becker A & Kabsch W (2002) X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA - how the enzyme uses the Cys-418 thiyl radical for pyruvate cleavage. *J Biol Chem* 277(42):40036-40042.

139. Andersson J, Bodevin S, Westman M, Sahlin M, & Sjoberg BM (2001) Two active site asparagines are essential for the reaction mechanism of the class III anaerobic ribonucleotide reductase from bacteriophage T4. *J Biol Chem* 276(44):40457-40463.

140. Wei Y*, et al.* (2014) A chemically competent thiosulfuranyl radical on the *Escherichia coli* class III ribonucleotide reductase. *J Am Chem Soc* 136(25):9001-9013.

141. Feliks M, Martins BM, & Ullmann GM (2013) Catalytic mechanism of the glycyl radical enzyme 4-hydroxyphenylacetate decarboxylase from continuum electrostatic and QC/MM calculations. *J Am Chem Soc* 135(39):14574-14585.

Chapter 2: Discovery and validation of an anaerobic microbial choline utilization gene cluster[a]

2.1: Introduction

As described in Chapter 1, choline and trimethylamine (TMA) are important nitrogen-containing metabolites that perform fundamental roles in biological pathways throughout nature. Choline is an essential nutrient for higher organisms, including humans, contributing to cell membrane function, methyl transfer events, and neurotransmission (1). The volatile odorant TMA is used as a carbon source by bacteria, is a precursor to the marine osmolyte trimethylamine *N*-oxide (TMAO), and is converted to the powerful greenhouse gas methane by methanogenic archaea (2-4). The sole biochemical reaction directly connecting these two small molecules is the metabolism of choline to TMA by anaerobic microorganisms (**Figure 2.1**). As detailed in Chapter 1, this chemical transformation plays a significant role in many biological systems and has a profound effect on human health.

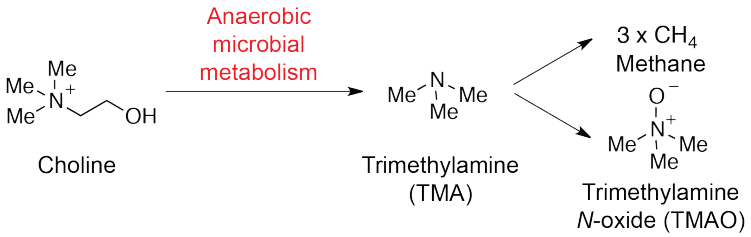


Figure 2.1: Generation of trimethylamine (TMA) from choline by anaerobic microbes and its subsequent processing.

Despite long-standing interest in this microbial metabolic pathway and its broad relevance to humans, nothing was known about its underlying genetics or biochemical mechanism prior to our work, as the enzymes responsible for anaerobic choline utilization had not been identified. This chapter describes our discovery of a gene cluster involved in anaerobic choline cleavage (*cut*, choline utilization cluster), and the identification of the enzymes responsible for TMA formation. Through multiple approaches, we demonstrate that the pathway encoded by the *cut* genes is present and functional in a

[a] This chapter is an unofficial adaptation of the following articles: **Craciun S**, Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109(52):21307–21312, and Martínez-del Campo A, **Bodea S**, Hamer HA, Marks JA, Haiser HJ, Turnbaugh PJ, Balskus EP (2015) Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *MBio* 6(2):e00042-15.

diverse range of human gut bacteria. We anticipate that our discovery and analyses of the *cut* gene cluster will aid efforts to further understand the role of choline metabolism in the human gut microbiota and its link to disease.

2.2: Results and discussion

2.2.1: Discovery of a choline utilization gene cluster in Desulfovibrio desulfuricans

Our strategy to discover a choline utilization gene cluster relied upon the existence of a sequenced bacterial species previously found to convert choline into TMA and acetaldehyde under anaerobic conditions (5), and the fact that bacteria often tend to co-localize genes encoding proteins involved in a particular metabolic pathway into gene clusters. Our search for candidate genes involved in microbial choline degradation was also guided by the hypothesis that some of the enzymatic transformations in this pathway resemble those used in catabolism of the structurally related metabolite ethanolamine (**Figure 2.2**). Bacteria from the human gut and other environments utilize ethanolamine as a carbon and nitrogen source by cleaving its C–N bond to afford acetaldehyde and ammonia. The proteins required for ethanolamine utilization are encoded by the *eut* gene cluster, which has been studied extensively in the past 50 years in pathogenic strains of *Salmonella enterica* and *Escherichia coli* (6). A study of a mouse model of intestinal infection showed that ethanolamine could provide a significant growth advantage to *Salmonella enterica* in the lumen of the inflamed intestine, when its utilization as a carbon source is coupled with the use of host-derived tetrathionate as a terminal electron acceptor (7). Biochemical studies have determined that the C–N bond cleavage of ethanolamine is performed by ethanolamine ammonia-lyase (EutBC), a vitamin B_{12}-dependent enzyme that generates ammonia and acetaldehyde (8). The acetaldehyde produced in this initial step is further processed by the alcohol dehydrogenase EutG and the aldehyde oxidoreductase EutE to give ethanol and acetyl coenzyme A (acetyl-CoA) (6). In addition to genes that encode for catabolic enzymes, *eut* clusters also typically contain genes encoding bacterial microcompartment proteins, which are known to assemble and form icosahedral organelle-like structures that encapsulate the entire metabolic pathway (9). It was proposed that the *eut* microcompartment might function to facilitate the sequestration of the volatile acetaldehyde

intermediate in a manner analogous to concentration of carbon dioxide by the carboxysome, or to protect the cell from the potential toxicity of acetaldehyde (10).

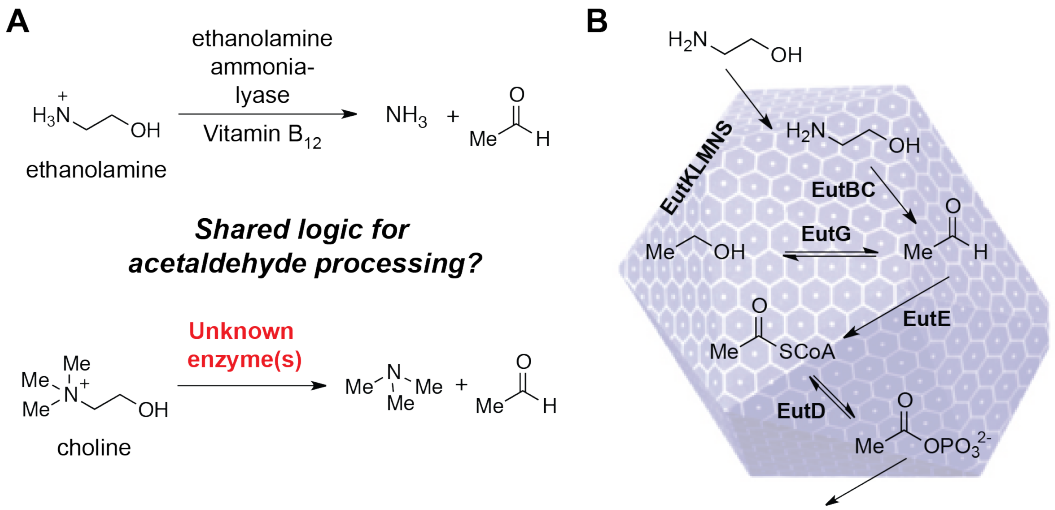


Figure 2.2: Anaerobic microbial degradation of ethanolamine. **(A)** Potential parallel logic between anaerobic choline utilization and bacterial ethanolamine utilization pathways. **(B)** Model for ethanolamine catabolism within a protein microcompartment. Microcompartment structural proteins: EutK, EutL, EutM, EutN and EutS; ethanolamine ammonia lyase: EutBC; alcohol dehydrogenase: EutG; acetaldehyde dehydrogenase, EutE; phosphotransacetylase: EutD.

Recognizing that acetaldehyde is formed in the initial step of both ethanolamine and choline degradation pathways (8), we hypothesized that its conversion to downstream products might be similar. We used position-specific iterative-basic local alignment search tool (PSI-BLAST) to search for genes encoding homologs of EutG, EutE, and microcompartment protein EutM from *Salmonella enterica* in the genome of *Desulfovibrio desulfuricans* American Type Culture Collection (ATCC) 27774. This sulfate-reducing bacterium had been previously reported to metabolize choline to TMA (5), a finding that we independently reconfirmed by growing the bacteria in media containing (trimethyl-d_9)-choline (d_9-choline) (**Figure 2.3**). Our search revealed homologs of all three *eut* genes clustered within a 16.6-kb genomic region that displays similar organization to the *eut* cluster (**Figure 2.4, Table 2.1**). We named this portion of the *D. desulfuricans* genome the choline utilization (*cut*) gene cluster. Remarkably, homologs of *eutB* and *eutC*, which encode ethanolamine ammonia-lyase, were absent from both this 19-gene cluster and the rest of the *D. desulfuricans* genome. Instead, the *cut* cluster contained a distinct set of

genes predicted to encode a glycyl radical enzyme CutC (Ddes_1357) and a glycyl radical activating protein CutD (Ddes_1358).

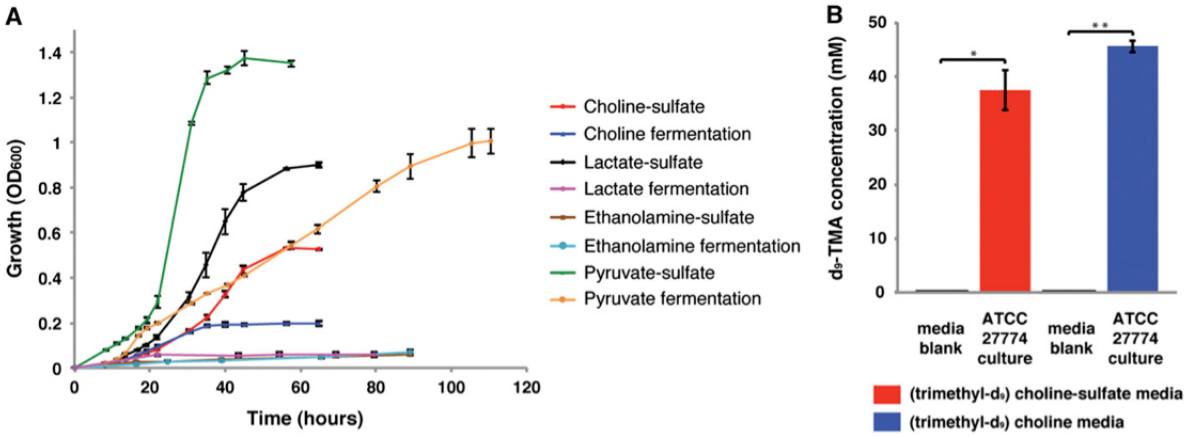


Figure 2.3: Choline utilization by *Desulfovibrio desulfuricans* ATCC 27774 proceeds via formation of trimethylamine (TMA). **(A)** Growth of *D. desulfuricans* ATCC 27774 using various carbon sources. The data shown are the average OD_{600} values of two cultures. Error bars represent standard deviation (SD). **(B)** LC-MS quantification of d_9-TMA production during *D. desulfuricans* ATCC 27774 growth on d_9-choline. Choline-sulfate and choline fermentation media contained 60 mM d_9-choline chloride. Bar graphs represent the mean ± standard error of mean (SEM) of four replicates. *P < 0.002; **P < 10^{-4}.

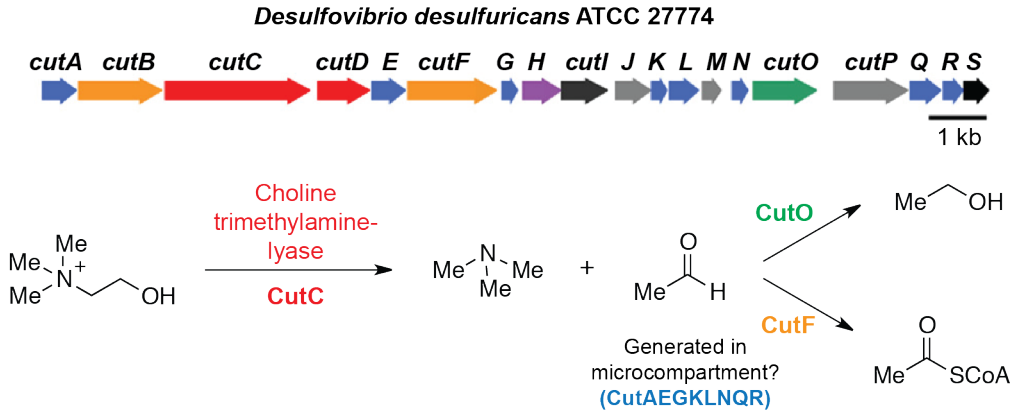


Figure 2.4: Putative choline utilization (*cut*) gene cluster and proposed biochemical pathway for microbial choline metabolism.

Table 2.1: Annotation of choline utilization (*cut*) gene cluster in *D. desulfuricans* ATCC 27774.

Gene	GenBank accession no. (protein sequence)	Size (bp/aa)	Homolog in Eut operon (% amino acid ID/similarity)	Putative function
cutA Ddes_1355	ACL49257	597/188	EutM (51/73) EutK (37/62)	Microcompartment structural protein
cutB Ddes_1356	ACL49258	1475/492	EutE (29/48)	CoA acylating aldehyde dehydrogenase
cutC Ddes_1357	ACL49259	2574/848	None	Glycyl radical enzyme
cutD Ddes_1358	ACL49260	933/310	None	Glycyl radical enzyme activating protein
cutE Ddes_1359	ACL49261	609/202	EutM (50/67) EutK (48/68)	Microcompartment structural protein
cutF Ddes_1360	ACL49262	1578/525	EutE (32/50)	CoA acylating aldehyde dehydrogenase
cutG Ddes_1361	ACL49263	288/99	EutM (71/83) EutK (47/71)	Microcompartment structural protein
cutH Ddes_1362	ACL49264	684/227	None	Propanediol utilization protein
cutI Ddes_1363	ACL49265	822/273	EutJ (43/62)	Putative chaperonin
cutJ Ddes_1364	ACL49266	639/212	None	Propanediol utilization protein
cutK Ddes_1365	ACL49267	288/99	EutN (38/61)	Microcompartment structural protein
cutL Ddes_1366	ACL49268	549/182	EutM (40/57) EutK (41/60)	Microcompartment structural protein
cutM Ddes_1367	ACL49269	336/111	None	Propanediol utilization protein
cutN Ddes_1368	ACL49270	285/94	EutM (71/84)	Microcompartment structural protein
cutO Ddes_1369	ACL49271	1107/368	EutG (36/51)	Alcohol dehydrogenase
cutP Ddes_1370	ACL49272	1326/441	None	NADH dehydrogenase
cutQ Ddes_1371	ACL49273	552/183	EutM (35/58)	Microcompartment structural protein
cutR Ddes_1372	ACL49274	348/115	EutS (52/72) EutM (34/57)	Microcompartment structural protein
cutS Ddes_1373	ACL49275	435/144	EutP (34/52)	Ethanolamine utilization protein

Bioinformatic analyses provided strong support for annotation of CutC as a member of the glycyl radical enzyme family. As described in Chapter 1, glycyl radical enzymes (GREs) are widespread in anaerobic microbes and use highly reactive protein-based radical intermediates to promote a diverse set of chemical transformations, including C–C bond formation (benzylsuccinate synthase), C–C bond cleavage (pyruvate formate-lyase, 4-hydroxyphenylacetate decarboxylase), and dehydration (glycerol dehydratase) (**Figure 2.5.A**). However, genome sequencing and bioinformatics have revealed numerous uncharacterized GREs within microbial genomes, suggesting that the functional diversity of these enzymes is underappreciated (11). We hypothesized that CutC might catalyze the initial TMA-forming step in choline metabolism, although C–N bond cleavage by glycyl radical enzymes has not previously been reported. Maximum likelihood phylogenetic analysis (12) of CutC, its close homologs, and biochemically characterized glycyl radical enzymes revealed that these sequences cluster according to function and that CutC and its homologs form a separate clade (**Figure 2.5.B**). Both alignment of the CutC amino acid sequence with those of functionally characterized glycyl radical enzymes (**Figure 2.6.A**) and construction of a homology model using Modeller (**Figure 2.6.B**) revealed conservation of key active site glycine and cysteine residues involved in radical catalysis (13). Multiple sequence alignments of the

predicted choline TMA-lyase homologs from the new GRE clade showed conservation of predicted active site residues, supporting the hypothesis of shared function (**Figure 2.6.C**). Taken together, these findings strongly suggested that CutC possesses a biochemical function distinct from known glycyl radical enzymes, that of a choline TMA-lyase.

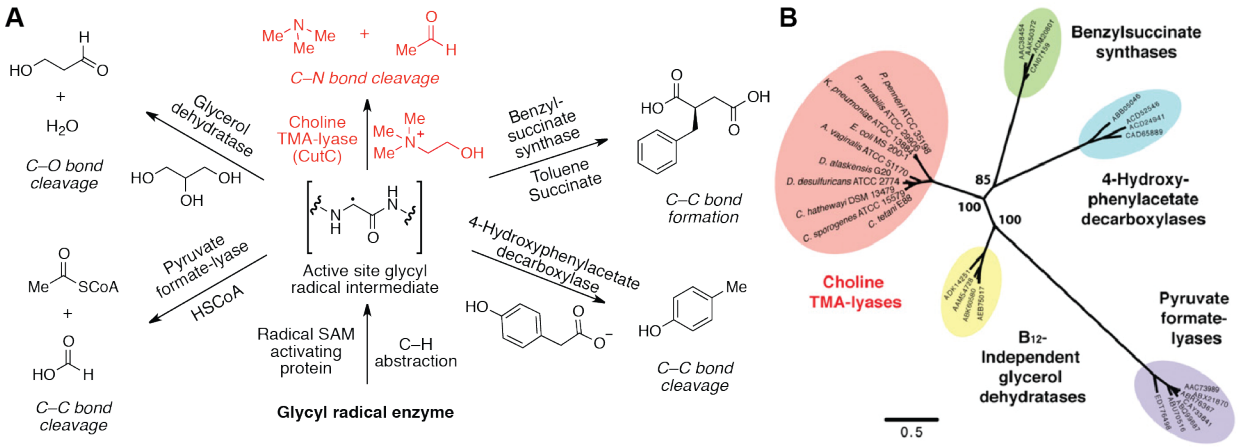


Figure 2.5: Bioinformatics identify choline trimethylamine-lyase CutC as a distinct type of glycyl radical enzyme. **(A)** Biochemical transformations catalyzed by glycyl radical enzymes. **(B)** Maximum likelihood phylogenetic tree showing the relationship between the amino acid sequences of choline TMA-lyases and other biochemically characterized members of the glycyl radical enzyme family. Bootstrap confidence values >50 are indicated on the nodes.

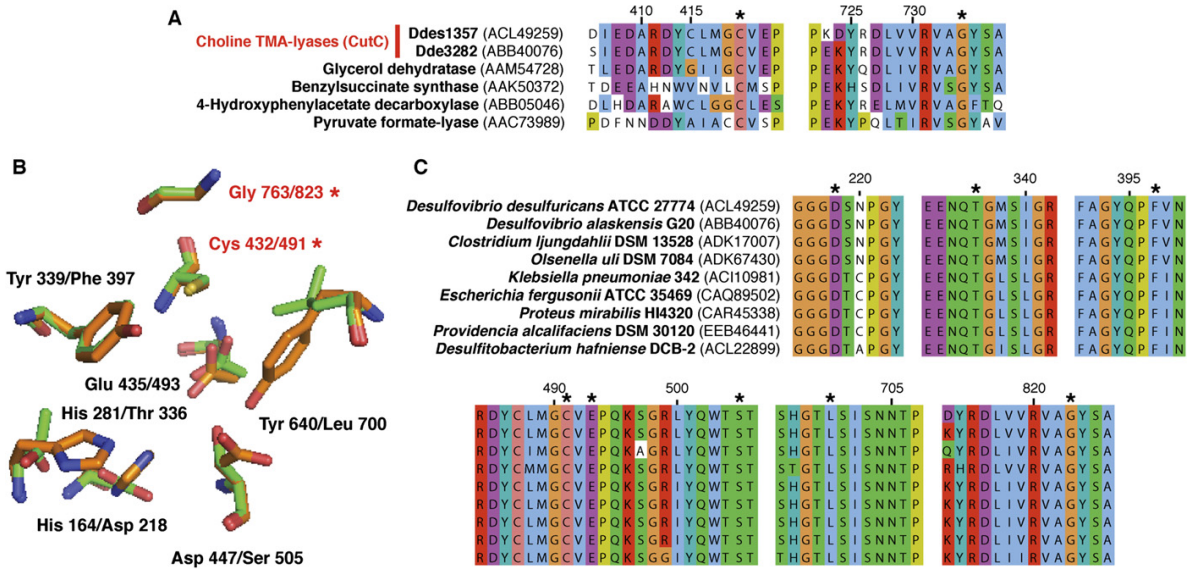


Figure 2.6: Bioinformatic analyses of choline TMA-lyase CutC. **(A)** Multiple sequence alignment of putative choline TMA-lyases from *Desulfovibrio* species with characterized glycyl radical enzymes reveals conservation of catalytic active site residues (shown with asterisks). *E. coli* pyruvate formate-lyase numbering is shown. GenBank accession numbers identify amino acid sequences used. **(B)** Overlay of the active site residues of B_{12}-independent glycerol dehydratase from *Clostridium butyricum* (crystal structure, orange) and choline TMA-lyase CutC (Ddes_1357) from *D. desulfuricans* (homology model, green). Asterisks indicate conserved putative active site residues of *D. desulfuricans*. **(C)** Multiple sequence alignment of putative choline TMA-lyases from different bacterial genera shows conservation of active site residues. Asterisks indicate conserved active site residues identified from the homology model shown in (B). *D. desulfuricans* ATCC 27774 CutC (Ddes_1357) numbering is shown.

2.2.2: Disruption of choline TMA-lyase CutC in Desulfovibrio alaskensis *G20 impairs growth on choline and abolishes production of TMA*

We devised two complementary approaches to evaluate whether the *cut* gene cluster was involved in choline utilization: genetic knockout in a choline-degrading organism and heterologous expression of the putative choline TMA-lyase CutC and its activating protein CutD in a non choline-using host. Among sequenced bacteria containing homologs of the putative choline degradation gene cluster was *Desulfovibrio alaskensis* G20, a sulfate-reducing bacterium capable of anaerobic choline utilization (14). A G20 mutant strain [C10(pB6)] with a disruption in the gene encoding predicted choline TMA-lyase CutC (Dde_3282) had been generated previously using transposon mutagenesis, but a mutant phenotype had not been clearly identified (15). We obtained both wild-type *D. alaskensis* G20 and strain C10(pB6) from the Krumholz group at University of Oklahoma and evaluated their ability to grow using lactate and choline as carbon sources (**Figure 2.7.A**). Consistent with earlier studies (15), we observed no

difference in growth between wild type and C10(pB6) using media containing lactate. In contrast, the two strains exhibited a marked discrepancy in their ability to grow on choline. Whereas the wild-type strain could both ferment choline and use it for growth with sulfate as a terminal electron acceptor, strain C10(pB6) was unable to grow in either choline-containing media. These results clearly connect CutC to anaerobic choline metabolism.

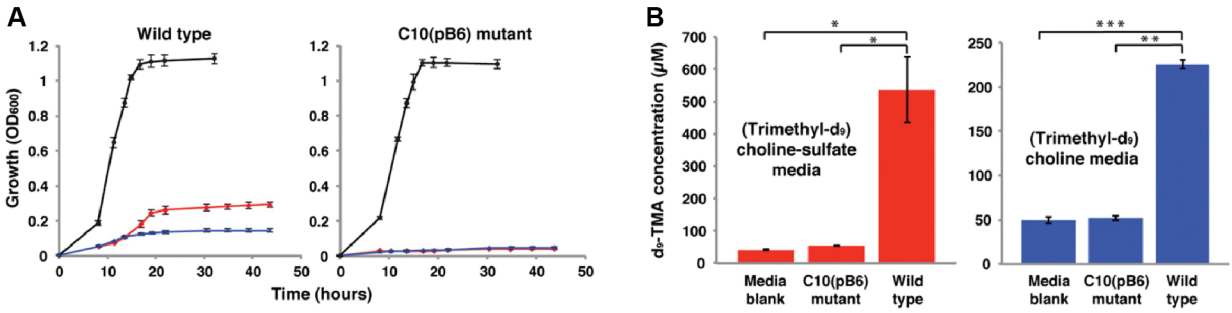


Figure 2.7: The *cut* gene cluster is responsible for choline metabolism and TMA production. **(A)** Growth of *D. alaskensis* G20 wild type (WT) and C10(pB6) strains at 37 °C on lactate sulfate (black), choline sulfate (red), and choline fermentation (blue) media. The data shown are the average OD_{600} values of four cultures. Error bars represent standard error of the mean (SEM). **(B)** LC-MS quantification of d_9-TMA produced by *D. alaskensis* G20 WT and C10(pB6) during incubations in d_9-choline sulfate and d_9-choline fermentation media containing d_9-choline chloride (60 mM). Bar graphs represent the mean ± SEM of four cultures. *P < 0.02; **P < 10^{-5}; ***P < 10^{-6}.

We performed further experiments to validate the involvement of the *cut* gene cluster in TMA production. To verify that choline utilization in *D. alaskensis* proceeds via formation of TMA, we used liquid chromatography-mass spectrometry (LC-MS) to measure the d_9-TMA present after inoculation and incubation of wild-type and C10(pB6) strains in media containing d_9-choline. Use of isotopically labeled substrate ensured quantification of only TMA derived directly from choline. Growth of the wild-type organism was accompanied by the formation of d_9-TMA, whereas levels for the C10(pB6) mutant did not exceed that of an uninoculated media control (**Figure 2.7.B**). Detection of TMA during growth on choline and the lack of TMA production from the mutant confirmed that the identified gene cluster encodes a pathway that processes choline via C–N bond cleavage.

2.2.3: Heterologous expression of choline TMA-lyase in Escherichia coli

In parallel with these analyses we evaluated the in vivo activity of the putative TMA-forming enzymes CutC and CutD using a heterologous expression strategy. A 3.4-kb fragment of the *D. alaskensis*

G20 gene cluster encoding both choline TMA-lyase CutC (Dde_3282) and its activating protein CutD (Dde_3281) was cloned into a pET29b(+) vector and transformed into *Escherichia coli* BL21(D3). Using LC-MS, we measured formation of d_9-TMA from d_9-choline in vivo by strains harboring either the choline-degrading enzymes or an empty expression vector. Only *E. coli* coexpressing CutC and CutD produced d_9-TMA at levels above an uninoculated media control (**Figure 2.8**), verifying that these enzymes are responsible for the initial C–N bond cleavage in anaerobic choline degradation. Expression of either enzyme on its own completely abolished TMA production, which is consistent with their annotations as a glycyl radical enzyme and activating protein; generation of the key active site glycyl radical involved in catalysis by glycyl radical enzymes requires the action of a separate radical *S*-adenosylmethionine (SAM) enzyme (**Figure 2.5.A**) (16). Site-directed mutagenesis experiments with CutC provided further support for a radical-based mechanism of C–N cleavage, as mutation of either Gly821 or Cys489 to Ala prevented TMA formation. These universally conserved active site residues are the locations of the two protein-based radicals involved in catalysis and are essential for glycyl radical enzyme function (13).

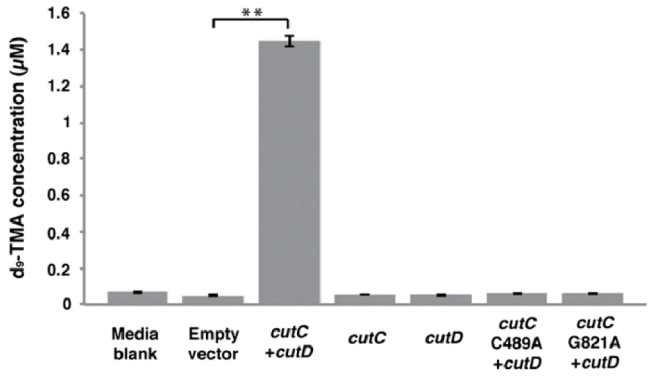


Figure 2.8: CutC and CutD are responsible for choline cleavage. LC-MS quantification of d_9-TMA produced by *E. coli* BL21(DE3) during incubations in Luria–Bertani (LB) medium containing d_9-choline chloride (100 μM). Bar graphs represent the mean ± SEM of four cultures. **$P < 10^{-5}$.

2.2.4: Growth of Desulfovibrio desulfuricans *on choline involves a glycyl radical enzyme*

We obtained additional evidence connecting a glycyl radical enzyme to choline metabolism using electron paramagnetic resonance (EPR) spectroscopy. We prepared cell suspensions of *D. desulfuricans* ATCC 27774 grown on either choline or pyruvate fermentation media and recorded their EPR spectra

(**Figure 2.9**). The characteristic signal of a glycine-centered radical was observed only for cells grown on choline. The g value (2.003) and strong hyperfine coupling to a single proton (1.3 mT) indicate the presence of an organic radical located at the α-carbon of an amino acid and are consistent with previously reported EPR spectra of cells expressing glycyl radical enzymes (17, 18). The weaker nonhyperfine isotropic signal observed for cells from pyruvate cultures likely originates from flavoproteins, which are produced by a related *Desulfovibrio* strain during growth on pyruvate (19). Overall, these EPR experiments link choline utilization to a glycine-centered radical and indicate that choline TMA-lyase CutC is a member of the glycyl radical enzyme family.

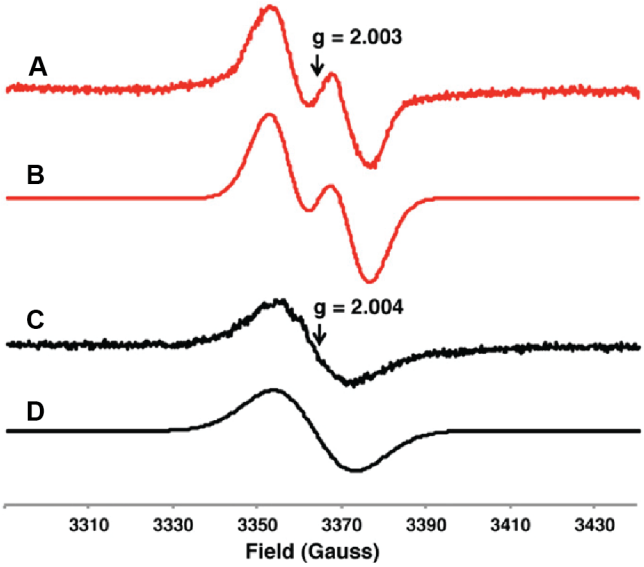


Figure 2.9: Growth of *D. desulfuricans* on choline involves a glycyl radical. EPR spectra of *D. desulfuricans* ATCC 27774 cell suspensions grown on choline (red) and pyruvate (black) fermentation media. **(A)** Experimental spectrum of cells grown on choline. **(B)** Simulation assuming an isotropic signal with g = 2.003, line width (W) = 1.14 mT, and isotropic hyperfine interaction with one proton (A) = 1.3 mT. **(C)** Experimental spectrum of cells grown on pyruvate. **(D)** Simulation assuming an isotropic signal with g = 2.004 and line width (W) = 1.94 mT.

2.2.5: Bioinformatic analysis of the distribution of the cut *gene cluster in sequenced bacterial genomes*

Our discovery of a choline utilization gene cluster has implications beyond elucidating the biochemical basis for TMA formation; it also greatly enhances our understanding of the phylogenetic and environmental distribution of this metabolic activity by ascribing function to previously uncharacterized or misannotated gene sequences. Together with a postdoctoral fellow in the Balskus group, Dr. Ana

Martínez-del Campo, I performed searches for CutC homologs in November 2014 against the nonredundant protein database in the National Center for Biotechnology Information (NCBI) and against all genomes in the IMG database. The results revealed *cut* gene clusters in a total of 459 bacterial genomes (88% to 61% amino acid identity) distributed across four bacterial phyla (*Proteobacteria*, *Firmicutes*, *Actinobacteria*, and *Fusobacteria*). The majority of the hits were found in *Proteobacteria* (*Gammaproteobacteria* and *Deltaproteobacteria*) and *Firmicutes* (*Clostridia* and *Bacilli*), with less than 3% from *Actinobacteria* (*Coriobacteridae*) and *Fusobacteria* (*Fusobacteriia*). CutC was conspicuously absent from *Bacteroidetes*, a major phylum of the gut microbiota (20). All CutC-containing organisms were facultative or obligate anaerobes, which is consistent with the oxygen sensitivity of GREs (13). CutC-encoding strains originated from a range of microbial habitats, including the human body (50% of hits), with many of the human isolates obtained from the gut (35%). This finding strengthened earlier hypotheses that anaerobic choline degradation is a major source of TMA formation within this environment (21) and provided initial insights regarding which symbiotic microbes may harbor this metabolic activity. We also identified *cutC* in three genomes assembled from human gut metagenomic DNA (*Clostridium hathewayi* CAG:224 [PRJNA221953], *Collinsella* sp. CAG:289 [PRJNA222147], and *Klebsiella variicola* CAG:634 [PRJNA221971]). *CutC* was also observed in gut isolates from other animals, including insects, fish, and mammals. Other strains harboring a copy of *cutC* were isolated from different parts of the human body (urogenital tract, oral cavity, and airways), as well as from the environment. The environmental microbes were largely found in marine sediments, habitats that are rich in TMA-based methanogenesis, confirming that choline metabolism contributes to TMA production in these ecosystems. Overall, we observed a discontinuous distribution of the *cut* gene cluster across most microbial species (**Figure 2.10**). For example, only 4% (98) of the 2,719 *Escherichia coli* genomes and 15% (16) of the 101 *Streptococcus suis* genomes encode a CutC homolog. *Desulfosporosinus* (6/6 genomes) and *Proteus* (17/17 genomes) are the only two bacterial genera that universally harbor the *cut* cluster. Other species that frequently harbor *cutC* include *Clostridium botulinum* (41/62 genomes; 66%) and *Pectobacterium* (18/29 genomes; 62%). These results suggest that anaerobic choline utilization is a

widely distributed pathway that may be important for survival in many microbial habitats. In addition to providing insights from genomes of cultured microbes, the ability to identify the choline utilization pathway in sequencing data will also permit characterization of its role in natural microbial communities through use of culture-independent methods such as metagenomics and metatranscriptomics (22).

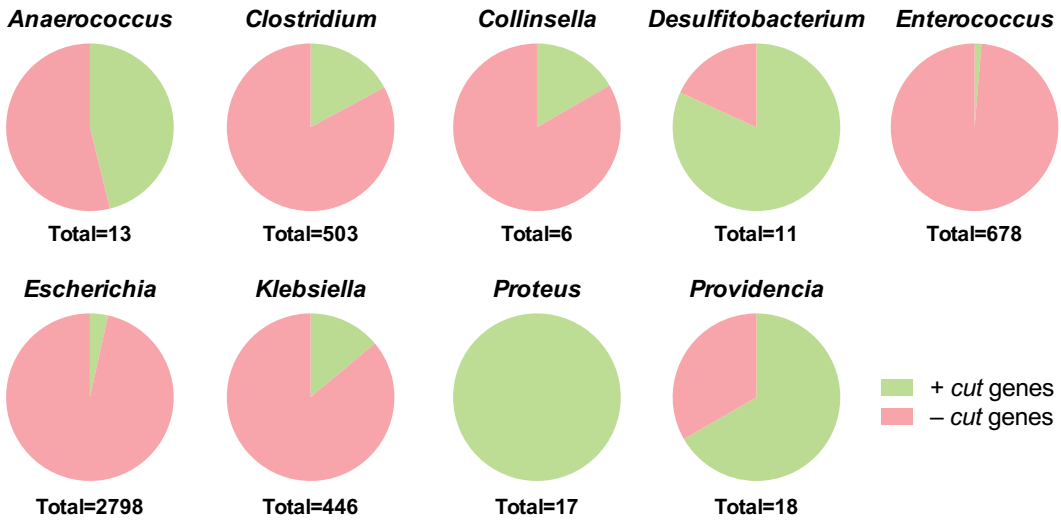


Figure 2.10: The *cut* gene cluster is widely but unevenly distributed in sequenced human gut bacteria. The pie charts depict the proportions of bacterial genomes that contain the *cut* gene cluster (green) in selected bacterial genera that possess this pathway and have been isolated from the human gut.

We examined the genomic context of the *cutC* homologs and found that the additional elements of the *D. desulfuricans cut* gene cluster are largely conserved in other bacteria. The majority of the *cutC* sequences (~95%) clustered with homologs of 10 or more *D. desulfuricans cut* genes. A subset of these *cut* clusters (~10%) lacked genes encoding the acetaldehyde oxidoreductase, alcohol dehydrogenase, and/or microcompartment structural proteins, but in all cases homologs of the missing enzymes were identified elsewhere in the genome. *CutC* and *cutD* homologs failed to cluster with any other *cut* genes in only 3 of the 459 genomes, and only one strain possessed *cutC* but lacked *cutD* (*Olsenella* sp. oral taxon 809 F0356). We classified *cut* clusters into two groups (type I and type II), based on their gene content (**Figure 2.11.A**). Type I clusters, found in *Firmicutes*, *Actinobacteria*, and *Deltaproteobacteria*, contained ~20 genes, including homologs of BMC-encoding genes (*cutAEGKLNQR*), *cutB*, *cutC*, *cutD*, *cutF*, *cutHIJ*, *cutO*, *cutS*, and in some cases *cutP*. Although gene content was consistent for these clusters, gene order and transcriptional orientation varied. In contrast to the type I clusters, type II *cut* clusters, found

exclusively in facultative anaerobes from *Gammaproteobacteria*, typically contained only 10 genes, including the central enzymes involved in choline fermentation (*cutC*, *cutD*, *cutF*, *cutO*, and *cutH*) and four BMC structural proteins. The CutC homologs from these clusters possess an additional N-terminal domain and are longer by ~300 amino acids than homologs from type I clusters. There is a possibility that the differences in the arrangement and distribution of *cut* clusters and CutC sequences have arisen in response to various exposures of choline-metabolizing organisms to molecular oxygen. Further experiments will need to be carried out to verify this hypothesis.

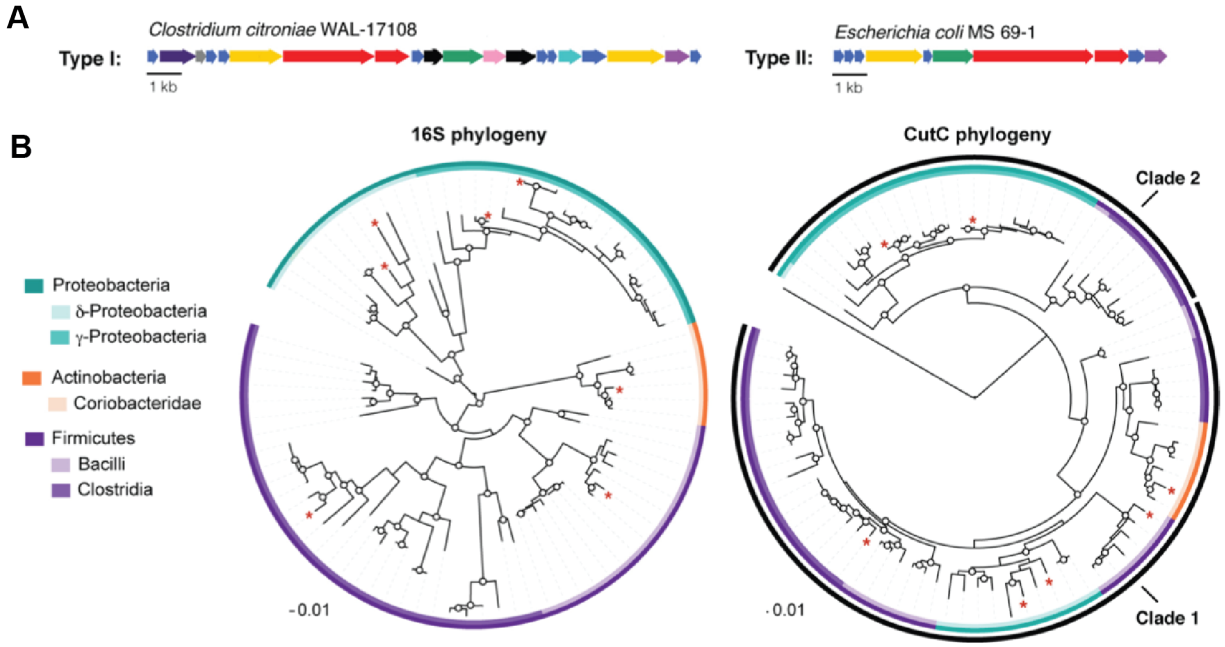


Figure 2.11: (A) Representative type I and type II *cut* gene cluster arrangements. **(B)** Maximum-likelihood phylogenetic trees showing the relationship between 16S rRNA genes (left) and the relationship between CutC amino acid sequences (right). Bootstrap confidence values of >50 are indicated by circles on the nodes. The CutC tree also contains the amino acid sequence of the glycyl radical enzyme glycerol dehydratase as an outgroup. Sequences from experimentally characterized choline-metabolizing strains are indicated with a red asterisk.

With the exception of the extra N-terminal domain found in CutC homologs from type II clusters, we found that the amino acid sequence of this key TMA-generating enzyme is highly conserved across organisms possessing the *cut* cluster (88% to 61% amino acid identity). To deduce the phylogenetic relationship between choline TMA-lyases from different organisms, Dr. Martínez-del Campo, constructed a maximum-likelihood phylogenetic tree of CutC amino acid sequences and compared it to the phylogeny of the 16S rRNA gene (**Figure 2.11.B**). She observed two distinct clades of CutC sequences, one

consisting of sequences from *Actinobacteria*, *Deltaproteobacteria*, and *Firmicutes* (clade 1) and the other containing sequences from *Gammaproteobacteria* and *Peptococcaceae* (clade 2). Notably, this arrangement is incongruent with the 16S rRNA-based phylogeny, suggesting that *cutC* has been acquired through horizontal gene transfer. This hypothesis is also supported by the presence of putative mobile genetic elements in the vicinity of 35% of the *cut* clusters, including the uropathogenic *E. coli* strain 536, whose cluster is located within a known pathogenicity island (PAII) (23). We propose that acquisition of *cut* genes might occur in habitats in which choline is an abundant growth substrate, as it would confer a competitive advantage in such environments. Recent work has suggested that horizontal gene transfer occurs more frequently in the human gut than in other habitats, potentially due to the high population densities and increased metabolic activity in this setting (24). While exchange of antibiotic resistance genes among gut microbes has received particular attention (25), there is growing evidence for transfer of genes encoding metabolic pathways (26, 27). The *cut* gene cluster therefore represents an additional example of a bacterial metabolic function that may have been exchanged in the gut environment.

2.2.6: The cut *gene cluster is a genetic marker for anaerobic choline metabolism*

To confirm that the *cut* gene cluster is a functional marker for anaerobic choline metabolism, Dr. Martínez-del Campo and I tested whether its presence in a genome was predictive of an organism's ability to convert choline to TMA under anaerobic conditions. We obtained nine sequenced strains containing type I *cut* clusers (*Clostridium citroniae* WAL-17108, *Streptococcus dysgalactiae* subsp. *equisimilis* ATCC 12394, *Olsenella* sp. oral taxon 809 F0356, *Olsenella uli* DSM 7084) and type II *cut* clusters (*Escherichia coli* MS 69-1, *Klebsiella* sp. MS 92-3, *Proteus mirabilis* ATCC 29906, *P. mirabilis* HI4320, and *P. mirabilis* BB2000). These organisms represented each of the phyla identified in our bioinformatics search and, with the exception of *S. dysgalactiae*, were all human isolates. Each strain was grown anaerobically in rich medium supplemented with 1 mM d_9-choline to approximate choline availability in the human gut (28). All organisms generated d_9-TMA except for one of the *Olsenella* strains (**Figure 2.12)**. This strain, *Olsenella* sp. oral taxon 809 F0356, was missing the GRE activase *cutD*, whereas the other *Olsenella* strain, *Olsenella uli* DSM 7084, contained both *cutC* and *cutD*. The failure of the strain

harboring an incomplete *cut* cluster to metabolize choline provided further evidence that activated CutC is required for this activity. Interestingly, all of the strains metabolized choline in rich media containing alternate carbon and energy sources. This observation suggests that choline is a preferred substrate or that this pathway contributes to cellular processes distinct from growth. Deciphering the specific advantages associated with this mode of metabolism will require further experiments, but we hypothesize that preferential use of choline is indicative of resource partitioning, providing bacteria harboring this cluster with a unique metabolic niche and allowing them to avoid competition for more energy-rich growth substrates available in the gut environment (29).

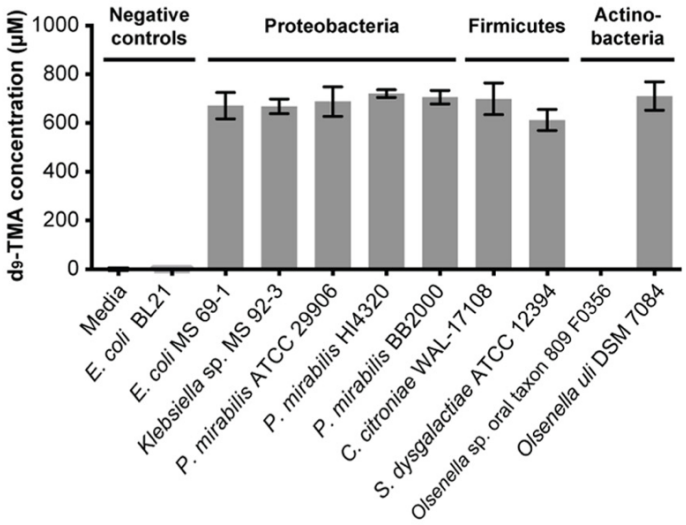


Figure 2.12: The *cutC* functional gene is diagnostic of anaerobic choline utilization. LC-MS quantification of d_9-TMA produced by selected *cutC*-containing bacterial species grown in rich medium supplemented with d_9-choline (1 mM); *Olsenella* sp. oral taxon 809 F0356 has *cutC* but is missing the gene encoding activase *cutD*. Each bar represents the mean ± standard deviation (SD) of d_9-TMA from three cultures.

2.2.7: Development and validation of a PCR method for detection of choline TMA-lyase

Understanding the role of anaerobic choline metabolism in the human gut and its connection to disease requires a means of identifying and quantifying this activity in individual gut communities. The discovery that bacterial phylogeny is not a good marker for choline utilization led us to explore methods for direct detection of the functional gene *cutC*. We began these efforts by developing a PCR method that could selectively amplify *cutC* in the presence of additional genes encoding functionally diverse GREs

possessing similar sequences (13). Our primer design took advantage of our understanding of the relationship between CutC's amino acid sequence and catalytic activity derived from the homology model from **Figure 2.6**. By aligning 85 sequences, we identified conserved stretches of amino acids not found in other GREs and designed over 20 potential pairs of degenerate PCR primers targeting these regions. We screened each of these candidate primer pairs for their ability to amplify *cutC* from genomic DNA isolated from a subset of sequenced strains with *cut* gene clusters. One of these primer pairs displayed particularly robust amplification across a range of strains and was chosen for further optimization. These primers, which target regions of the *cutC* gene that encode conserved active site residues (Phe395, Cys489, and Glu491 from *D. alaskensis* G20 CutC), amplify a 314-bp portion of the gene that is readily distinguished from the corresponding regions of genes encoding unrelated GREs.

We validated the utility of this degenerate PCR approach by successfully amplifying *cutC* from the genomic DNA of 11 strains containing *cut* clusters. A product of the expected size was observed in all of the PCRs, with amplification efficiency varying moderately among the different isolates (**Figure 2.13.A**). Dr. Martínez-del Campo confirmed the identity of each PCR product through cloning and sequencing; in all cases, *cutC* was the only sequence obtained. No amplification was observed with DNA from a strain lacking a *cut* cluster but possessing other GREs (*Roseburia inulinovorans* A2-194). Because these PCR primers selectively amplified *cutC* from phylogenetically diverse isolates, they should be useful for surveying strains from different taxonomic groups. Finally, Dr. Martínez-del Campo performed serial dilutions to establish that our primers can detect as little as ~10 pg of *cutC*-containing genomic DNA template, even in the presence of an excess of genomic DNA from an organism lacking this gene.

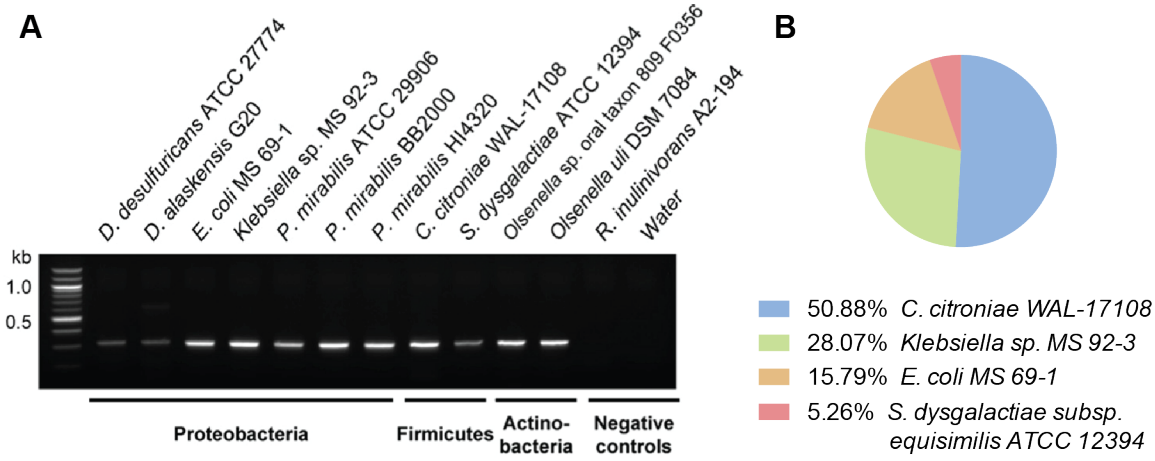


Figure 2.13: The functional gene *cutC* can be detected using degenerate PCR. **(A)** Agarose gel electrophoresis results for degenerate PCRs using *cutC*-specific degenerate primers and genomic DNA from representative *cutC*-containing, choline-utilizing strains and two sequenced organisms that lack *cutC*. **(B)** Recovery of *cutC* sequences from a mixed DNA sample by applying degenerate primers. The sample contained equal amounts of DNA from four different *cut*-cluster containing isolates.

We also utilized the degenerate primers to amplify *cutC* sequences directly from human gut metagenomic DNA. Dr. Martínez-del Campo first assessed whether biases in amplification efficiency might impact recovery of *cutC* sequences from a mixed DNA template by applying the primers to a mock community containing equal amounts of template DNA from four different isolates possessing the *cut* cluster, including a strain that had displayed lower levels of amplification in our original survey (*S. dysgalactiae*). Successful amplification of *cutC* sequences from 4 ng of community template DNA was followed by construction and sequencing of a 57-member clone library. All of the sequences present in the original sample were recovered, although in different proportions (**Figure 2.13.B**). This promising result prompted us to perform a proof-of-concept degenerate PCR with metagenomic DNA isolated from a human fecal sample that possessed the ability to convert choline to TMA ex vivo (**Figure 2.14.A**). We observed a band of the appropriate size in the PCR product from this complex sample and constructed a clone library of the amplified products. Sequencing of 66 clones revealed 33 unique *cutC* nucleotide sequences encoding 23 distinct amino acid sequences with high identity to CutC proteins from seven sequenced bacteria spanning three different genera (**Figure 2.15.B**). This experiment provided a first glimpse into the diversity of *cutC* sequences present in an individual human gut microbial community,

revealing that multiple types of bacteria have the capacity to perform anaerobic choline metabolism in this environment. We note that while the translated *cutC* sequences obtained in this experiment matched CutC protein sequences from sequenced bacteria, many of these sequences did not show significant similarity to known *cutC* nucleotide sequences, suggesting we have accessed sequences from organisms that are not yet represented in genome sequence databases.

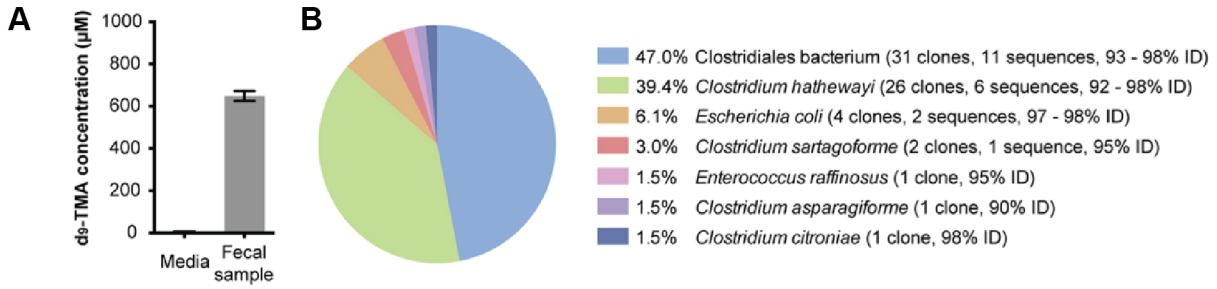


Figure 2.14: (A) LC-MS quantification of d_9-TMA produced ex vivo by a human fecal sample upon anaerobic incubation in BHI medium supplemented with (trimethyl-d_9)-choline (1 mM) for 18 h at 37°C. Bar graphs represent the means ± SD of three independent incubations. **(B)** Abundances of translated *cutC* sequences in the 66-member clone library constructed from degenerate PCR with human gut metagenomic DNA and the percent identity to CutC proteins from sequenced isolates.

Studies of particular metabolic pathways in microbial habitats are greatly facilitated by the use of molecular approaches that target functional genes, particularly when the activities of interest have a broad phylogenetic distribution. Previous efforts to characterize gut microbial metabolism using function-oriented strategies had largely focused on pathways and activities that are well understood in comparison to anaerobic choline metabolism, including bile acid metabolism, short-chain fatty acid synthesis, and glucuronidase activity (30-32). These studies have provided important insights into the distribution and abundance of these core functions in gut microbial communities, including information about how activities and pathways vary between individuals and how functional gene abundance changes with diet. Our work illustrates the benefits of applying such approaches, which are greatly enabled by the wealth of available bacterial genome sequences, to study newly characterized functions.

2.2.8: Isolation and characterization of new choline-utilizing human gut isolates

An important unanswered question regarding the pathway encoded by the *cut* gene cluster is the extent to which it represents a predominant route for choline utilization and TMA generation in the

human gut. We explored this problem by screening new choline-metabolizing isolates from this environment for the presence of the *cutC* functional gene. We obtained isolates from two human fecal samples via enrichment culturing on six different media formulations that contained choline as a single carbon source (**Figure 2.15.A**). We thus aimed to isolate only bacteria that could utilize choline for growth and minimize the possibility of other media components being used as carbon sources. We included either ammonium, Casitone, or Trypticase peptone as the nitrogen source, and also varied the electron acceptors present (nitrate, tetrathionate, fumarate, or none). Organisms were chosen for further characterization based on their ability to convert d_9-choline to d_9-TMA in either brain heart infusion (BHI) or enrichment medium (**Figure 2.15.B, C**). Using this approach, we accessed a total of nine distinct choline-utilizing strains from four genera (*Clostridium*, *Klebsiella*, *Escherichia*, and *Enterococcus*) as determined by 16S rRNA gene sequencing. We obtained *cutC* sequences from each of the isolates by using degenerate PCR (**Figure 2.15.D**), and analyzed the phylogeny of each translated *cutC* sequence (**Figure 2.15.E**). This experiment revealed that our new choline-utilizing strains universally possess the *cutC* gene. To our knowledge, our approach represents the first targeted isolation of choline-utilizing bacteria from the human gut. Importantly, we were able to correlate anaerobic choline metabolism in each isolate with the presence of the *cutC* gene. These results further establish the validity of the *cutC* gene as a molecular marker for anaerobic choline metabolism and demonstrate the utility of directly detecting this biochemical function. Most significantly, the observation that these gut isolates universally possess *cutC* confirms that choline-utilizing organisms are present in human gut microbial communities, and suggests that the pathway encoded by the *cut* cluster could be the dominant mechanism for choline degradation and TMA production within this environment.

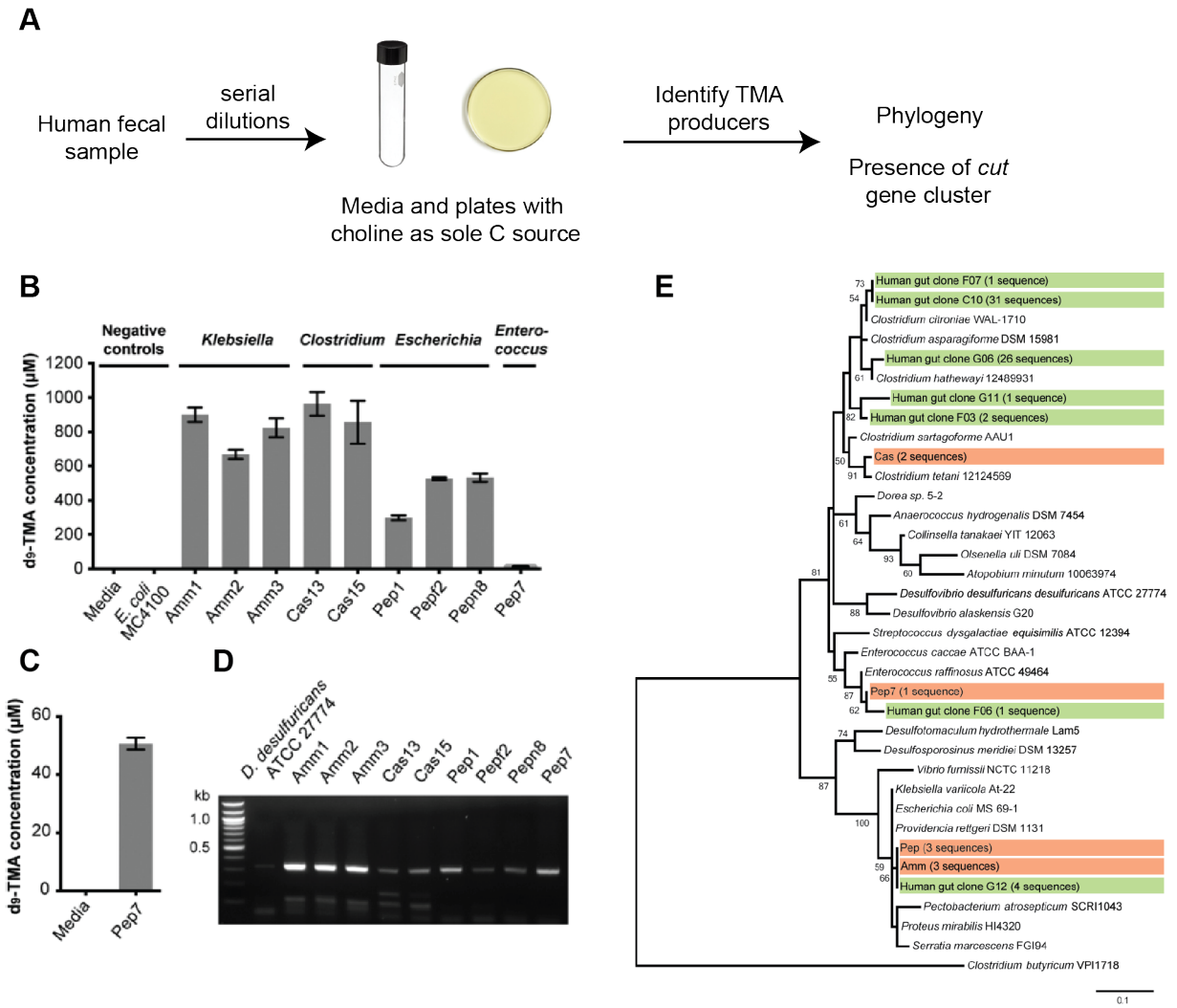


Figure 2.15: Choline-utilizing isolates from the human gut universally possess *cutC*. **(A)** General strategy for enrichment culturing. **(B)** LC-MS quantification of d_9-TMA produced by human gut isolates grown anaerobically in BHI medium supplemented with d_9-choline (1 mM) for 20 h at 37°C. Bar graphs represent the means ± standard deviations (SD) of three independent cultures. **(C)** LC-MS quantification of d_9-TMA produced by isolate Pep7 grown anaerobically in enrichment medium supplemented with d_9-choline (1 mM) for 20 h at 37°C. Bar graphs represent the means ± SD of three independent cultures. **(D)** Agarose gel electrophoresis results for PCRs with *cutC*-specific degenerate primers and genomic DNA isolated from choline-utilizing human isolates. **(E)** Phylogenetic analysis of the *cutC* sequences from the choline-utilizing strains accessed through enrichment culturing.

2.2.9: Conclusions

The search for genes involved in bacterial choline utilization has led to the discovery of an additional member of the glycyl radical enzyme family, choline TMA-lyase. At the outset of our studies, it was not obvious what type of enzyme would be involved in producing TMA from choline. From the standpoint of biochemical logic, the initial C–N bond cleavage event in choline utilization is most similar

to catabolic pathways involving vitamin B_{12}-dependent radical rearrangements (ethanolamine and propanediol utilization) (33). However, no gene cluster encoding a vitamin B_{12}-dependent enzyme has been linked to choline degradation and no enzyme from this family has ever been shown to catalyze this transformation. A critical distinction between choline and metabolites processed by B_{12}-dependent pathways, such as ethanolamine, is the presence of a trimethylammonium substituent. Choline lacks N–H bonds and therefore cannot make all of the hydrogen bonding contacts within an enzyme-active site that are considered important for catalysis of B_{12}-dependent eliminations (8, 33). Although this crucial structural difference called into question whether analogies in pathway logic between choline and ethanolamine utilization would extend to the C–N bond cleaving event, it was evident that powerful enzymatic chemistry would be needed to perform this transformation. The annotation of CutC as a glycyl radical enzyme provided a potential solution to supplying the requisite catalytic activity for C–N bond scission.

Large-scale microbial genome sequencing has afforded an unprecedented opportunity to uncover the genetic basis for important activities associated with complex microbial communities. Our discovery of the genes involved in anaerobic choline utilization illustrates the power of combining biochemical knowledge with bioinformatics to correlate sequence data with function. These findings immediately increased our understanding of this metabolic pathway's distribution in biology and expanded the scope of known radical-mediated enzymatic chemistry. Ultimately, our work will enable a better understanding of anaerobic choline metabolism, both in the environment, where it impacts the global carbon cycle and greenhouse gas production, and within the human body, where it contributes to disease-associated interspecies metabolism.

Our preliminary experimental and bioinformatic findings suggest that the pathway encoded by the *cut* gene cluster may be a major mechanism for the direct conversion of choline to TMA in gut bacteria. However, many central questions remain regarding the factors influencing TMA production by the human gut microbiota and TMAO levels in humans. Choline may be utilized via alternate pathways that do not yield TMA, such as oxidation to betaine (34) and demethylation to produce dimethylaminoethanol (35).

The factors that influence levels of choline and other potential TMA precursors in the gut, including diet, availability of other bacterial metabolic activities such as lipid metabolism, and other environmental factors, are not well established. For example, TMA and TMAO levels may also be influenced by the presence of TMA-utilizing archaea (36). We expect that the ability to directly detect choline TMA-lyase in gut microbial communities, along with other "omics"-based approaches, will prove broadly useful in future efforts to correlate this metabolic pathway with metabolite levels, to identify alternate routes for choline utilization, and to understand the connection between this bacterial activity and human disease. More generally, our findings illuminate the importance of employing methods that directly detect the microbial functional genes responsible for activities of interest. We anticipate that implementation of similar function-oriented approaches will help reveal gut microbial processes that are the bases for disease and can potentially guide therapeutic development.

2.3: Materials and methods

2.3.1: Materials and general methods

All chemicals were purchased from Sigma-Aldrich, except for d_9-trimethylamine (TMA) hydrochloride (CDN Isotopes), and (trimethyl-d_9)-choline chloride (Cambridge Isotope Laboratories). Methanol, acetonitrile and water used for liquid chromatography-mass spectrometry (LC-MS) were B&J Brand high-purity solvents (Honeywell Burdick & Jackson). Oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA). Recombinant plasmid DNA was purified with a QIAGEN Plasmid Mini Kit (Qiagen, Valencia, CA). Purification of PCR reactions and gel extraction of DNA fragments for restriction endonuclease clean-up were performed using an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) or a QIAquick Gel Extraction Kit (Qiagen). DNA sequencing was performed by Beckman Coulter Genomics (Danvers, MA) or by GENEWIZ (Boston, MA). SDS/PAGE gels were purchased from Bio-Rad. Optical densities of *Desulfovibrio desulfuricans*, *Desulfovibrio alaskensis*, and *Escherichia coli* cultures were determined with a GENESYS 20 spectrophotometer (Thermo Scientific) or with a DU 730 Life Sciences UV/Vis spectrophotometer (Beckman Coulter, Brea, CA) by measuring absorbance at 600 nm.

LC-MS/MS analysis was performed in the Small Molecule Mass Spectrometry Facility at Harvard University on an Agilent 6460 Triple Quadrupole Mass Spectrometer with Agilent 1290 uHPLC (Agilent Technologies, Wilmington, DE) (for derivatized d_9-TMA detection) or in our research labs in the Department of Chemistry and Chemical Biology, Harvard University, on an Agilent 6410 Triple Quadrupole LC/MS instrument (Agilent Technologies, Wilmington, DE) (for derivatized and underivatized TMA and d_9-TMA). Samples and blanks were introduced via an electrospray ionization (ESI) source. The mass spectrometers were operated in multiple reaction monitoring (MRM) mode. The capillary voltage was set to 4.0 kV and the fragmentor voltage to 110 V (underivatized and derivatized TMA and d_9-TMA). The drying gas temperature was maintained at 350 °C (Agilent 6460) or 200 °C (Agilent 6410) with a flow rate of 12 L/min (Agilent 6460) or 10 L/min (Agilent 6460) and a nebulizer pressure of 25 psi (derivatized TMA and d_9-TMA) or 45 psi (underivatized d_9-TMA). The precursor-product ion pairs used in MRM mode were: m/z 69.1→m/z 51 (underivatized d_9-TMA), m/z 155.1→ m/z 66.2 (derivatized d_9-TMA), m/z 146.1→ m/z 58.1 (derivatized TMA). The collision energies for the precursor-product ion pairs were 21 V (underivatized d_9-TMA) and 43 V (derivatized TMA and d_9-TMA). MS1 resolution was set to wide, MS2 resolution was set to unit, the time filter width used was 0.07 min, and the ΔEMV (Electron Multiplier Voltage) was 400 V. Data analysis was performed with Mass Hunter Workstation Data Acquisition software (Agilent Technologies). The LC analysis for derivatized TMA and d_9-TMA was performed in positive ion mode, using a Bio-Bond (Dikma Technologies) C4 column (5 μm, 4.6 mm × 50 mm), preceded by a C4 guard column (3.5 μm, 2.0 mm × 20 mm). The mobile phase consisted of a 50/50 water/methanol mixture, supplemented with 5 mM ammonium formate and 0.1% formic acid as solvent modifiers. The flow rate was maintained at 0.3 mL/min for each run. Run time per sample was 6 min, and the first 1.8 min of flow through was diverted to waste. The injection volume was 3 μL. Blanks consisting of 50/50/0.025 (vol/vol/vol) acetonitrile/water/ formic acid were run in between each sample. For underivatized d_9-TMA, the LC analysis was performed in positive mode, using a Kinetex (Phenomenex) HILIC column (2.6 μm, 30 mm × 2.1 mm, 100 Å), preceded by a C4 precolumn (3.5 μm, 2.0 mm × 20 mm). The LC conditions were:

80% B for 1.6 min, a gradient decreasing to 40% B over 2.4 min, 40% B for 2 min, a gradient increasing to 80% B over 2 min, and 80% B for 3 min (solvent A = 5 mM ammonium formate in water with 0.1% formic acid, solvent B = 100% acetonitrile). The flow rate was maintained at 0.3 mL/min for each run. Run time per sample was 9 min, and the first 1.4 min of flow through was sent to waste. The injection volume was 3 μL for both standards and samples. For quantification of d_9-TMA, an external standard curve was prepared in BHI media supplemented with 1 mM choline-d_9 and analyzed at concentrations ranging from 104 pg/μL to 627 pg/μL d_9-TMA. To assess the statistical significance of individual data points, a Student t test (two-tailed, heteroscedastic) was used to calculate a P value testing the null hypothesis that the amount of d_9-TMA is the same in the samples indicated on each graph.

Perpendicular mode X-band electron paramagnetic resonance (EPR) spectra were recorded on a Bruker ElexSysE500 EPR instrument with a 100 K–600 K Digital Temperature Control system, ER 4131VT (Bruker BioSpin), and data acquisition and baseline corrections were performed with Xepr software (Bruker). The magnetic field was calibrated with a standard sample of α,γ-bisdiphenylene-β-phenylallyl (BDPA), g = 2.0026 (Bruker). The experimental spectra were modeled with EasySpin for Matlab to obtain g values, hyperfine coupling constants, and line widths (37). EPR spectra represent the average of nine scans and were recorded under non-saturating conditions: temperature, 120 K; center field, 3,360 Gauss; range, 150 Gauss; microwave power, 63.25 μW; microwave frequency, 9.433602 GHz (choline fermentation media), 9.436572 GHz (pyruvate fermentation media); modulation amplitude, 0.1 mT; modulation frequency, 100 kHz; time constant, 40.96 ms; conversion time, 163.84 ms; and scan time, 167.77 s.

2.3.2: Identification of the cut *(choline utilization) gene cluster in sequenced bacterial genomes*

A putative choline utilization gene cluster was identified within the sequenced genome of *D. desulfuricans* ATCC 27774 using the BLAST search tool at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Specifically, the following protein sequences from the ethanolamine utilization operon of *Salmonella enterica* subsp. *enterica serovar Typhimurium* were used as input for individual PSI-BLAST

searches of the *D. desulfuricans* ATCC 2774 genome: alcohol dehydrogenase EutG (AAA80211), aldehyde oxidoreductase EutE (AAA80209), and microcompartment protein cchA/EutM (AAA80207). The genes encoding the closest homologs of each of these three protein sequences were colocalized within the *D. desulfuricans* genome.

The choline TMA-lyase (CutC) protein sequence from *D. desulfuricans* ATCC 27774 (EMBL accession number ACL49259.1) was used as a query for BLASTP searches against the nonredundant protein sequences (nr) database from NCBI (performed on 27 November 2014) and all genomes in the IMG database (performed on 29 November 2014) in order to locate additional choline utilization gene clusters. BLAST hits having an identity between 88 and 61% to CutC from *D. desulfuricans* were aligned using ClustalW2 (38) to confirm the presence of conserved active site residues. Strains containing a *cutC* homolog were classified according to the source of the isolate based on the information available at the Genome Online Database, the IMG database, or bibliographic searches. The content and arrangement of each of the *cut* clusters identified in the initial BLAST search were determined by the examination of the genetic context of the *cutC* gene. Putative functions were assigned to neighboring genes based on the genome annotations and similarities to the *cut* genes from *D. desulfuricans*.

2.3.3: Bioinformatic analyses of choline TMA-lyase CutC

Nucleotide and protein sequences were aligned with ClustalW2 using the default parameters. A CutC (Ddes_1357) homology model was generated with the HHPred interactive server (http://toolkit.tuebingen.mpg.de/hhpred) (39) using the crystal structure of B_{12}-independent glycerol dehydratase (40) as a template [Protein DataBase (PDB) ID 1R9D]. An overlay of the CutC homology model with the glycerol dehydratase crystal structure revealed conservation of active site architecture and the essential catalytic glycine and cysteine residues. Multiple sequence alignment of Ddes_1357 with CutC homologs from other organisms containing *cut* gene clusters showed conservation of active site residues predicted from the CutC homology model. The maximum likelihood phylogenetic tree of representative glycyl radical enzymes was reconstructed using PHYML (12) with the general time reversible (GTR) nucleotide substitution model, the proportion of invariable sites and the γ parameter of

across-site rate variation (using four rate categories) was estimated from the dataset. Bootstrap support values were calculated with the same parameters (100 replicates). The sequence dataset used for this analysis was obtained starting from choline TMA-lyase Ddes_1357, other functionally characterized representatives of the glycyl radical enzyme family, and close BLAST hits from the GenBank nucleotide database (**Table 2.2**). These sequences were aligned using ClustalW and trimmed manually using MacClade. Translated CutC nucleotide sequences from isolates obtained via enrichment culturing and metagenomic clone libraries were aligned with CutC sequences from bacterial genomes. These alignments were manually edited to a final length of 104 aa. Maximum-likelihood trees were constructed in MEGA 5 (41) using the Jones-Taylor-Thorton (JTT) method of substitution for CutC and the Tamura-Nei method for 16S rRNA. Statistical support for both trees was obtained by bootstrapping 100 iterations. Trees were visualized with iTOL (42).

Table 2.2: Glycyl radical enzyme amino acid sequences used in CutC phylogenetic analysis.

Gene or locus tag	Organism	GenBank accession no. (protein sequence)	Annotation
Ddes_1357	*Desulfovibrio desulfuricans* ATCC 27774	ABB40078	Choline TMA-lyase
Dde_3282	*Desulfovibrio alaskensis* G20	ABB40076	Choline TMA-lyase
CLOSTHATH_02755	*Clostridium hathewayi* DSM 13479	EFC99034	Choline TMA-lyase
CLOSPO_02864	*Clostridium sporogenes* ATCC 15579	EDU36695	Choline TMA-lyase
CTC_01449	*Clostridium tetani* E88	AAO36007	Choline TMA-lyase
HMPREF0078_1374	*Anaerococcus vaginalis* ATCC 51170	EEU12078	Choline TMA-lyase
HMPREF9553_01535	*Escherichia coli* MS 200–1	EFJ62362	Choline TMA-lyase
HMPREF0484_5138	*Klebsiella pneumoniae subsp. rhinoscleromatis* ATCC 13884	EEW38822	Choline TMA-lyase
HMPREF0693_2863	*Proteus mirabilis* ATCC 29906	EEI47333	Choline TMA-lyase
PROPEN_00404	*Proteus penneri* ATCC 35198	EEG87333	Choline TMA-lyase
dhaB1	*Clostridium butyricum*	AAM54728	Glycerol dehydratase
pflB1	*Clostridium ljungdahlii* DSM 13528	ADK14251	Glycerol dehydratase
NT01CX_1220	*Clostridium novyi* NT	ABK60580	Glycerol dehydratase
CbC4_0337	*Clostridium botulinum* BKT015925	AEB75017	Glycerol dehydratase
tutD	*Thauera aromatica*	AAC38454	Benzylsuccinate synthase
bssA	*Azoarcus* sp. T	AAK50372	Benzylsuccinate synthase
Geob_2448	*Geobacter* sp. FRC-32	ACM20801	Benzylsuccinate synthase
bssA	*Aromatoleum aromaticum* EbN1	CAI07159	Benzylsuccinate synthase
csdB	*Clostridium scatologenes*	ABB05046	4-Hydroxyphenylacetate decarboxylase
CLH_1872	*Clostridium botulinum* E3 str. Alaska E43	ACD52546	4-Hydroxyphenylacetate decarboxylase
CLL_A1723	*Clostridium botulinum* B str. Eklund 17B	ACD24941	4-Hydroxyphenylacetate decarboxylase
hpdB	*Clostridium difficile* ATCC 9689	CAD65889	4-Hydroxyphenylacetate decarboxylase
pflB	*Escherichia coli* str. K-12 substr. MG1655	AAC73989	Pyruvate formate-lyase
SARI_01990	*Salmonella enterica subsp. arizonae*	ABX21870	Pyruvate formate-lyase
pflB	*Klebsiella pneumoniae subsp. pneumoniae* MGH 78578	ABR76367	Pyruvate formate-lyase
pfl1	*Photorhabdus luminescens subsp. laumondii* TTO1	CAY33841	Pyruvate formate-lyase
CGSHiGG_03495	*Haemophilus influenzae* PittGG	ABQ99687	Pyruvate formate-lyase
VIBHAR_01546	*Vibrio harveyi* ATCC BAA-1116	ABU70516	Pyruvate formate-lyase
pflB	*Clostridium butyricum* 5521	EDT76498	Pyruvate formate-lyase

2.3.4: Growth of Desulfovibrio *bacterial strains on choline and quantification of TMA production from choline*

Desulfovibrio alaskensis G20 wild-type and C10(pB6) mutant strains were obtained from Lee Krumholz (University of Oklahoma, Norman, OK). The G20 strain (43) is a spontaneously nalidixic acid-resistant derivative of strain G100A, a known choline degrader (14). The C10(pB6) mutant strain was generated previously using signature-tagged mutagenesis; this mutant contains a transposon insertion that disrupts the gene encoding predicted choline TMA-lyase CutC (Dde_3282) and was initially identified in a screen for *D. alaskensis* genes conferring sediment fitness (15). *D. desulfuricans* ATCC 27774 was obtained from João Nuno Carita (Universidade Nova de Lisboa, Lisbon). All strains were grown anaerobically at 37 °C on lactate-sulfate (LS), choline-sulfate (CS), and choline fermentation (C) media as described below. Kanamycin (1,050 μg/mL) was added to all cultures of the C10(pB6) mutant. Lactate-sulfate (LS) medium was used for routine cultivation and was prepared as previously described (44). All concentration values indicated correspond to final values. LS medium contains sodium lactate (60 mM, 1.2% wt/vol), sodium sulfate (50 mM), magnesium sulfate (8 mM), ammonium chloride (5 mM), HEPES pH 7.2 (25 mM), calcium chloride (0.6 mM), yeast extract (Becton, Dickinson and Company) (0.1%), trace mineral solution [nitrilotriacetic acid trisodium salt (0.62 mM), iron (II) sulfate heptahydrate (19 μM), manganese (II) chloride tetrahydrate (6.3 μM), cobalt (II) chloride hexahydrate (8.9 μM), copper (II) chloride dihydrate (1.5 μM), zinc (II) sulfate heptahydrate (9.1 μM), sodium molybdate dihydrate (0.5 μM), and sodium selenate (1.0 μM)], vitamin solution (3) [pyridoxine hydrochloride (96 nM), thiamine hydrochloride (30 nM), riboflavin (27 nM), calcium pantothenate (42 nM), lipoic acid (48 nM), *p*-aminobenzoic acid (73 nM), nicotinic acid (81 nM), vitamin B (7 nM), biotin (16 nM), folic acid (9 nM), and mercaptoethanesulfonic acid (141 nM)], and resazurin (2.5 μM). The pH of the medium was adjusted to 7.2. The medium was autoclaved for 5 min, sparged for 30 min with argon, dispensed into 18 × 150 mm modified Hungate tubes (Chemglass Life Sciences; CLS-4209-10) and capped with butyl rubber stoppers and aluminum seals (10 mL media per tube). The tubes were autoclaved for an additional 20 min. After cooling, anaerobic and sterile aqueous solutions of potassium

phosphate dibasic (2.2 mM), sodium bicarbonate (8 mM), and cysteine-HCl (0.04%) were added. Tubes of anaerobic media were stored under a headspace of argon. Lactate fermentation media (L) was prepared in a similar manner to the LS media except that sodium sulfate was replaced with sodium chloride (100 mM) and magnesium sulfate was replaced with magnesium chloride (8 mM). Choline-sulfate (CS) and choline fermentation (C) media were prepared in a similar manner to the LS and L media, except that it was supplemented with 0.01 mg/mL iron (III) ammonium citrate, and sodium lactate was replaced by choline chloride (60 mM), which was added as an anaerobic, filter-sterilized aqueous solution shortly before inoculation. Ethanolamine- and pyruvate-containing media [ethanolamine-sulfate (ES), ethanolamine fermentation (E), pyruvate-sulfate (PS), and pyruvate fermentation (P)] were prepared like CS and C media, except that choline was replaced with ethanolamine hydrochloride (60 mM) or sodium pyruvate (60 mM). For *D. desulfuricans* ATCC 27774 growth on LS, CS, C, PS, and P media, two 18 × 150 mm modified Hungate tubes containing 10 mL of medium and N_2 in the headspace were inoculated with 0.1 mL of an overnight LS starter culture (OD_{600} = 0.85) and incubated at 37 °C. OD_{600} was measured periodically until the cultures reached stationary phase. For *D. alaskensis* G20 wild-type and C10(pB6) mutant strains growth on LS, CS and C media, four 18 × 150 mm Hungate tubes containing 10 mL of medium were inoculated with 0.05 mL (LS media) or 0.1 mL (CS and C media) of an overnight LS starter culture (OD_{600} = 1.0–1.1) and incubated at 37 °C until reaching stationary phase (~48 h).

For quantitation of d_9-TMA formation, the strains were grown on CS and C media containing (trimethyl-d_9)-choline (60 mM). This substrate was added as an anaerobic, filter-sterilized aqueous solution shortly before inoculation. Kanamycin (1,050 μg/mL) was added to all cultures of the C10(pB6) mutant. Four 10-mL cultures of each media type (LS, CS, and C) were grown side by side at 37 °C until reaching stationary phase (~48 h). The concentration of d_9-TMA in culture media was determined using LC-MS after derivatization with ethyl bromoacetate using a published procedure (45). Briefly, to a 10-mL glass vial sealed with a 13 × 20 mm sleeve stopper (VWR) was added 100 μL of bacterial culture filtered through a 13 mm, 0.22 μm pore-size Acrodisc syringe filter with HT Tuffryn Membrane (Pall Life

Sciences), 100 µL of TMA (43.3 mM in water), 10 µL of concentrated ammonia (7 M in methanol), and 120 µL of ethyl bromoacetate (20 mg/mL in acetonitrile). This mixture was incubated at room temperature for 30 min and then quenched by adding 4 mL of infusion solution [acetonitrile/water/formic acid, 50/50/0.025 (vol/vol/vol)]. A 0.5 µL aliquot of the resulting mixture was further diluted 2,000-fold with infusion solution and analyzed by LC-MS (3 µL injection volume) using the parameters described in materials and general methods. For analysis of *D. alaskensis* cultures, the derivatization reaction contained 100 µL of 22 µM aqueous TMA solution for the C10(pB6) mutant samples and 100 µL of 2.2 mM aqueous TMA solution for the wild-type samples. The quenched derivatization reactions for the C10(pB6) mutant were analyzed without dilution, whereas for the wild-type samples, a 10 µL aliquot was diluted 100-fold before analysis. Media blanks consisted of 10 mL of uninoculated CS and C media containing (trimethyl-d_9)-choline (60 mM).

2.3.5: Cloning of CutC (Dde_3282) and CutD (Dde_3281) from D. alaskensis *G20 and construction of CutC C489A and G821A mutants*

Genomic DNA from *D. alaskensis* G20 was isolated from a 10-mL LS culture using the UltraClean Microbial DNA Isolation kit from MoBio Laboratories. The sequences of primers used for amplification of *cutC* (Dde_3282) and *cutD* (Dde_3281) genes are provided in **Table 2.3**. Primers were designed to amplify both of the individual genes as well as the 3529-bp region (3270700–3274229) encoding both CutC and CutD. CutC was PCR amplified from genomic DNA using forward primer Dde-3282-NdeI-start and reverse primer Dde-3282-XhoI-nostop. All PCR reactions (set up in triplicate) contained 25 µL of PfuTurbo Hotstart master mix (Stratagene), 2 µL of genomic DNA template, and 100 pmol of each primer in a total volume of 50 µL. Thermocycling was carried out in a C1000 Gradient Cycler (Bio-Rad) using the following parameters: denaturation for 1 min at 95°C, followed by 40 cycles of 30 s at 95 °C, 1 min at 68.4°C, 5 min at 70 °C, and a final extension time of 10 min at 70 °C. PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide staining, pooled, and purified. Identical conditions were used to amplify *cutD* (forward primer Dde-3281-NdeI-start + reverse

primer Dde-3281-XhoI-stop) as well as *cutC* + *cutD* (forward primer Dde-3282-NdeI-start + reverse primer Dde-3281-XhoI-stop).

Table 2.3: Primers used for cloning and site directed mutagenesis

Oligo	Sequence	Restriction site
Dde-3282-NdeI-start	5′ GCATCATATGGATCTCCAGGACTTTTCACATAAGC 3′	NdeI
Dde-3282-XhoI-nostop	5′ GATTCTCGAGGAAACCATGCAGCATGG 3′	XhoI
Dde-3281-NdeI-start	5′ GCATCATATGAGAACCGCAACACACAGAGACG 3′	NdeI
Dde-3281-XhoI-stop	5′ GATTCTCGAGTCAGTGGCGGATCACCGAAACC 3′	XhoI
Dde3282-QC-C489A-for	5′ GACTACTGCCTGATGGGTGCCGTGGAACCGCAG 3′	—
Dde3282-QC-C489A-rev	5′ CTGCGGTTCCACGGCACCCATCAGGCAGTAGTC 3′	—
Dde3282-QC-G821A-for	5′ GTGGTGCGCGTGGCCGCATACAGCGCCTTCTTC 3′	—
Dde3282-QC-G821A-rev	5′ GAAGAAGGCGCTGTATGCGGCCACGCGCACCAC 3′	—

Amplified fragments were digested with NdeI and XhoI (New England Biolabs, NEB) for 2.5 h at 37 °C. Digests contained 1 µL of water, 3 µL of NEB buffer 4 (10×), 3 µL of BSA (10×), 20 µL of PCR product, 1.5 µL of NdeI (20 units/µL), and 1.5 µL of XhoI (20 units/µL). Restriction digests were purified directly using agarose gel electrophoresis and the gel fragments were further purified using the Illustra GFX kit. The digests were ligated into linearized expression vectors using T4 DNA ligase (NEB). The CutC + CutD and CutD digests were ligated into linearized pET-29b(+) to encode untagged and C-terminal His_6-tagged constructs, respectively, whereas the CutC digest was ligated into linearized pET-28a(+) to encode an N-terminal His_6-tagged construct. Ligations were incubated at 16 °C for 24 h and contained 3.5 µL water, 1 µL T4 ligase buffer (10×), 0.5 µL digested vector, 3 µL digested insert DNA, and 2 µL T4 DNA ligase (400 units/µL). A total of 5 µL of each ligation was used to transform a single tube of chemically competent *E. coli* TOP10 cells (Invitrogen). The identities of the resulting pET-29b-CutC/CutD, pET-29b-CutD, and pET-28a-CutC constructs were confirmed by sequencing purified plasmid DNA. These constructs were transformed into chemically competent *E. coli* BL21 (DE3) cells (Invitrogen) and stored at –80 °C as frozen LB/glycerol stocks.

The C489A and G821A mutants of CutC (Dde_3282) were constructed separately in pET-29b-CutC/CutD using site directed mutagenesis. The CutC (Ddes_1357) homology model identified these amino acid positions (which correspond to C491 and G823 in Ddes_1357 and are conserved in all glycyl radical enzymes) as sites of the thiyl and glycyl radicals involved in catalysis (11). The sequences of

primers used for mutagenesis are found in Table S5. The C489A mutant was constructed via PCR amplification of pET-29b-CutC/CutD using forward primer Dde3282-QC-C489A- for and reverse primer Dde3282-QC-C489A-rev. The G821A mutant was constructed via PCR amplification of pET-29b-CutC/CutD using forward primer Dde3282-QC-G821A-for and reverse primer Dde3282-QC-G821A-rev. PCR reactions contained 36 µL of PfuTurbo DNA polymerase (2.5 units/µL; Stratagene), 5 µL of Pfu DNA polymerase reaction buffer (10×), 1 µL of 10 mM dNTP Mix (Bio-Rad), 50 ng DNA template, and 125 ng of each forward and reverse primer in a total volume of 50 µL. Thermocycling was carried out using the following parameters: denaturation for 30 s at 95 °C, followed by 18 cycles of 30 s at 95 °C, 5 min at 55 °C, and 9 min at 68 °C. Upon cooling to room temperature, 1 µL of DpnI (NEB) was added to each tube, and the resulting mixtures were incubated at 37 °C. After 1 h, an additional 1 µL of DpnI was added to each tube and incubation continued for an additional hour at 37 °C. A total of 5 µL of each digestion was used to transform a single tube of chemically competent *E. coli* TOP10 cells. The identity of the resulting pET-29b-CutC/CutD-C489A or -G821A constructs was confirmed by sequencing purified plasmid DNA. Chemically competent *E. coli* BL21(DE3) cells were transformed with pure pET-29b-CutC/CutD-C489A or -G821A plasmid DNA and stored at –80 °C as frozen LB/glycerol stocks.

2.3.6: Heterologous expression of CutC and CutD and quantification of TMA production from choline

Individual cultures of LB medium (5 mL) containing 50 µg/mL kanamycin were inoculated with the pET-29b-CutC/CutD, pET-29b-CutD, pET-28a-CutC, pET-29b-CutC/CutD-C489A, and pET-29b-CutC/CutD-G821A expression strains, as well as a control strain of *E. coli* BL21(DE3) containing an empty pET-29b(+) vector, and incubated at 37 °C overnight. Aliquots (500 µL) of each overnight culture (OD_{600} = 1.6–1.8), were used to inoculate four 18 × 150 mm Hungate tubes containing 10 mL of LB medium (sparged with argon for 25 min) supplemented with NaCl (0.4 M), kanamycin (50 µg/mL), (trimethyl-d_9)-choline chloride (105 µM), and iron (III) ammonium citrate (0.069 mg/mL). High salt conditions were used to ensure expression of high-affinity choline transporter, *betT*, which is required for transport of choline into *E. coli* (46). The cultures were incubated at 37 °C with shaking (190 rpm), and

protein expression was induced with 500 μM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of ~0.3. Incubation continued for an additional 16 h at 37 °C. Four media blanks consisting of 18 × 150 mm modified Hungate tubes containing the growth media described above were incubated at 37 °C alongside the *E. coli* cultures. The concentrations of d_9-TMA in the media blanks and growth media of *E. coli* cultures were determined using LC-MS as described in materials and general methods. The transition used for quantitation was m/z 155.1→ 66.2. The derivatization reaction mixture contained the same components, except that the standard solution of TMA was not included. Quantitation was performed using a standard curve. Calibration standards were prepared by adding different concentrations of d_9-TMA hydrochloride solution to uninoculated LB media supplemented with 0.4 M NaCl (0.4 M), kanamycin (50 μg/mL), choline chloride (100 μM), and iron (III) ammonium citrate (0.066 mg/mL), such that the concentration of the derivatized d_9-TMA in the solution obtained after quenching would be in the 5 to 50 nM range. To ensure that the C498A and G821A CutC mutants were produced to the same extent as wild type, cell lysates from the heterologous expression experiments detailed above were analyzed using SDS-PAGE (**Figure 2.16**).

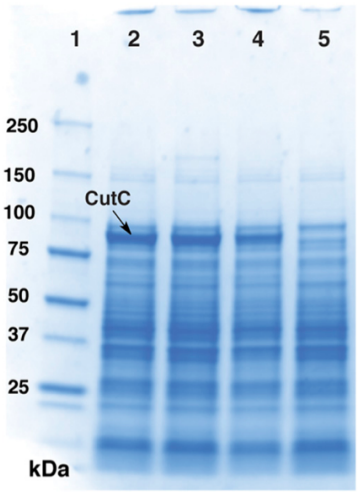


Figure 2.16: SDS-PAGE of cell lysates from heterologous expressions of CutC and CutD in *E. coli* using a 4–15% (wt/vol) polyacrylamide Tris-HCl gel. Lane 1, 10–250 kDa protein ladder (New England Biolabs); lane 2, wild-type CutC + CutD; lane 3, CutC C498A mutant + CutD; lane 4, CutC G821A mutant + CutD; and lane 5, empty pET-29b vector control. The calculated molecular weight of CutC (Dde_3282) is 96 kDa.

2.3.7: Whole cell electron paramagnetic resonance spectroscopy

D. desulfuricans ATCC 27774 was grown in serum bottles (outer diameter, 51.7 mm; height, 94.5 mm, Supelco; 33110-U) containing 100 mL of choline fermentation (C) and pyruvate fermentation (P) media using a 1 mL inoculum from an overnight LS culture (OD_{600} = 0.84–0.85). A total culture volume of 500 mL was used to prepare each cell suspension. The cultures were incubated at 37 °C in an anaerobic chamber (Coy Laboratory Products) under an atmosphere of 98% nitrogen and 2% hydrogen until reaching mid-to-late exponential phase. The cultures were combined in 2 × 250 mL polypropylene centrifuge tubes with plug seal caps (Corning; 430776) and centrifuged at 6,766 × g for 20 min at 4 °C. Cell pellets were resuspended in 0.1 M 3-(N-morpholino) propanesulfonic acid (MOPS) buffer (pH 7.2) containing 5 mM DTT to a final volume of 220–300 µL in the anaerobic chamber. The final volume of cell suspension was scaled according to the final average OD of the bacterial cultures. The cell suspensions were transferred to 4-mm thin-wall Suprasil screw-cap EPR sample tubes (7" length, Wilmad-LabGlass; 734-TR-7). The contents of the tubes were frozen in an acetone/dry-ice bath and placed in the cavity of the EPR instrument at 120 K. EPR spectra were recorded and modeled as described in materials and general methods.

2.3.8: Quantification of TMA production from choline for the other bacterial strains

To confirm and quantify the TMA formed during microbial choline utilization, reference strains were grown anaerobically on rich media containing 1 mM (trimethyl-d_9)-choline. LB media containing 20 mM fumarate was used for *E. coli* BL21 and MS 69-1 (BEI Resources, NIAID, Human Microbiome Project reference strain) and *Klebsiella sp.* MS 92-3 (BEI Resources, NIAID, Human Microbiome Project reference strain); BHI media was used for *P. mirabilis* ATCC 29906, *P. mirabilis* HI4320, *P. mirabilis* BB2000, *C. citroniae* WAL-17108 (BEI Resources, NIAID, Human Microbiome Project reference strain), and *S. dysgalactiae* subsp. *equisimilis* ATCC 12394; and PYG media was used for *Olsenella sp.* oral taxon 809 F0356 (BEI Resources, NIAID, Human Microbiome Project reference strain) and *Olsenella uli* DSM 7084 (DSMZ). Media was autoclaved for 5 min, sparged with nitrogen for 30 min, and dispensed under anaerobic conditions into 25 x 44 mm 5 mL serum vials (Supelco, Bellefonte, PA) or

18 x 160 mm modified Hungate tubes. Vials and tubes were subsequently capped with butyl rubber stoppers and aluminum seals (4 mL media per vial) and autoclaved for an additional 15 min. Three 4 mL replicates were inoculated with 40 μL of an overnight starter culture and grown for 20 h (10 days for *Olsenella* strains) at 37 °C. The concentration of d_9-TMA in culture media was determined using LC-MS after derivatization with ethyl bromoacetate. Media blanks consisted of 4 mL uninoculated media containing 1 mM (trimethyl-d_9)-choline.

2.3.9: Ex vivo incubation of a fecal sample with (trimethyl-d_9)-choline

A frozen fecal sample from a healthy individual was diluted 1:10 (wt/vol) in sterile prereduced phosphate-buffered saline (PBS). The dilution was vortexed for 1 min and allowed to settle for 15 min at room temperature. A 40-μl aliquot of the resulting supernatant was transferred to a capped 5-ml serum vial (Supelco, Bellefonte, PA) containing 4 ml of anaerobic BHI medium supplemented with 1 mM (trimethyl-d_9)-choline. Samples were incubated in triplicate for 18 h at 37 °C. The concentration of d_9-TMA in culture medium was determined using LC-MS after derivatization with ethyl bromoacetate. Medium blanks consisted of 4 ml uninoculated medium containing 1 mM (trimethyl-d_9)-choline.

2.3.10: Detection of cutC via degenerate PCR

To amplify an internal fragment of *cutC*, degenerate primers were designed based on an alignment from 85 CutC sequences from different species. These primers amplify a 314 bp conserved portion of the *cutC* gene. Primer sequences are 5'-TTYGCIGGITAYCARCCNTT-3' (Fw cutC-dPCR-389 aa-A) and 5'-TGNGGRTCIACYCAICCCAT-3 (Rv cutC-dPCR-492 aa-B). This primer set was used to amplify *cutC* from genomic DNA isolated from both reference strains and isolates obtained from enrichment culturing. PCR reactions contained 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 4 ng of DNA template, 0.2 μM of each primer, and 0.4 U of Takara Taq Hot Start Version (Takara Bio Inc., Otsu, Japan) in a total volume of 20 μL. Thermocycling was carried out in a MyCycler gradient cycler (Bio-Rad) with an initial denaturation period of 5 min at 95 °C, followed by 65 cycles of 30 s at 95 °C, 30 s at 58 °C, and 30 s at 72 °C, with a final extension time of 10 min at 72 °C. PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide staining. Bands of the expected length (314 bp) were

gel purified and cloned into the pCR4-TOPO vector using the TOPO TA Cloning Kit for Sequencing (Invitrogen). The identity of each clone was confirmed by sequencing.

To assess the amplification bias of the degenerate primers a 'mock' community DNA template containing equal amounts of genomic DNA from *E. coli* MS 69-1, *Klebsiella sp.* MS 92-3, *C. citroniae* WAL-17108, and *S. dysgalactiae* subsp. *equisimilis* ATCC 12394 was prepared. PCR reactions and cycling conditions were the same as described above. Amplified fragments were gel purified and cloned into the pCR4-TOPO vector using the TOPO TA Cloning Kit for Sequencing (Invitrogen). 60 clones were randomly chosen for analysis. Plasmid purification and sequencing were performed by Beckman Coulter Genomics (Danvers, MA). Low-quality, unsuccessful, and no-insert reactions were discarded reducing the number of clones to 57. Clones were identified by aligning *cutC* sequences from the corresponding organisms with the obtained nucleotide sequences.

2.3.11: Clone library construction and sequence analysis

A clone library specific for *cutC* was constructed by amplification of human fecal DNA with the degenerate primers described above. The PCR mixture and cycling parameters tested were the same as those used for the amplification of *cutC* from genomic DNA, except for the template DNA concentration, which was changed to 300 or 50 ng. PCR products were analyzed by agarose gel electrophoresis with ethidium bromide staining. Amplified PCR products from different template DNA concentrations were pooled, gel purified, and cloned into the pCR4-TOPO vector using the TOPO TA cloning kit for sequencing (Invitrogen) according to instructions from the manufacturer. A total of 96 clones were randomly selected for sequencing. Plasmid purification and sequencing were performed by Beckman Coulter Genomics (Danvers, MA). Low-quality, unsuccessful, and no-insert reaction mixtures were discarded, which reduced the size of the library to 66 clones. Nucleotide sequences were subjected to BLASTN searches against the NCBI and IMG databases and translated to protein sequences to perform a BLASTP search against the NCBI database. All translated sequences had 90 to 98% amino acid identity to CutC sequences in the database.

2.3.12: Enrichment culturing

Six types of media were prepared for enrichment culturing. All indicated concentration values correspond to final values. “A” media contained choline chloride (1 mM), ammonium chloride (20 mM), yeast extract (Becton, Dickinson and Company, Franklin Lakes, NJ) (0.1% w/v), potassium phosphate monobasic (3.3 mM), potassium phosphate dibasic (2.6 mM), sodium chloride (154 mM), magnesium chloride (0.4 mM), calcium chloride (0.8 mM), hemin (15 μM), sodium bicarbonate (4.8 mM), iron (III) ammonium citrate (0.1 mg/mL), HEPES pH 7.2 (25 mM), mineral solution (2), vitamin solution (2), and resazurin (2.5 μM). “C” media had the same composition as “A” media, except that ammonium chloride was replaced with casitone (Becton, Dickinson and Company) (1% w/v). “P” media contained choline chloride (60 mM), trypticase peptone (Becton, Dickinson and Company, Franklin Lakes, NJ) (1% w/v), yeast extract (Becton, Dickinson and Company, Franklin Lakes, NJ) (0.1% w/v), potassium phosphate monobasic (3.3 mM), potassium phosphate dibasic (2.6 mM), sodium chloride (154 mM), magnesium chloride (0.4 mM), calcium chloride (0.8 mM), hemin (15 μM), sodium bicarbonate (4.8 mM), iron (III) ammonium citrate (0.1 mg/mL), HEPES pH 7.2 (25 mM), mineral solution (2), vitamin solution (2), resazurin (2.5 μM), vitamin B_{12} (200 nM), vitamin K1 (0.0001% w/v) and L-cysteine (0.05% w/v). The pH of the media was adjusted to 7.2. The media was autoclaved for 5 min, sparged for 30 min with argon or nitrogen, dispensed into 18 × 150 mm modified Hungate tubes (Chemglass Life Sciences, Vineland, NJ; CLS-4209-10), and capped with butyl rubber stoppers and aluminum seals (10 mL media per tube). Tubes were autoclaved for an additional 20 min. Additionally; different electron acceptors were supplemented to “P” media by adding to each tube an anoxic sterile solution of fumarate, nitrate, or tetrathionate at a final concentration of 0.4 mM.

All experiments were performed under the guidance of the Harvard Committee on the Use of Human Subjects in Research. Two healthy volunteers (samples 1 and 2) provided an intact fecal sample that was stored at –80 °C before use. We do not have information regarding the diet and past antibiotic usage of the volunteers.

For the first type of enrichment ~200 mg of fecal sample 1 was thawed inside a Coy anaerobic chamber, dissolved in 1 mL of anoxic "A" media, and briefly centrifuged to allow particulates to settle. The mixture was diluted 10-fold with "A" media, and a 50 μL aliquot of each was used to inoculate tubes of "A" media and "C" media. The cultures were incubated for approximately 4 days at 37 °C before transferring to 10 mL of fresh media. The incubation and transfer steps were then repeated. At the end of each passage, a small volume (10 μL) of culture was plated on the corresponding agar media and incubated for 4 days at 37 °C. TMA levels in each of the liquid cultures were analyzed by LC-MS after derivatization with ethylbromoacetate. Qualitatively, the levels of TMA observed in the enrichment cultures were at least 1000-fold higher than in media blanks. At the end of the incubation period for the plates, 24 isolated colonies were chosen (3 from each plate in order to maximize the diversity of colony morphology) and used to inoculate tubes of corresponding media. The cultures were incubated for approximately 3 days at 37 °C, and their TMA levels were analyzed by LC-MS at the end of the incubation, as described above. Eight out of 24 cultures produced significant (greater than 1000-fold) amounts of TMA as compared to the media blanks.

In the second type of enrichment ~360 mg of frozen fecal sample 2 was diluted 1:10 (w/v) in pre-reduced PBS inside a Coy anaerobic chamber. Diluted samples were vortexed and large particulate matter was allowed to settle. A 100 μL aliquot was used to inoculate 10 mL of "P" media with no electron acceptor, fumarate, nitrate, or tetrathionate. The enrichments were incubated at 37 °C for two days and then transferred to fresh medium. After a two day incubation at 37 °C, cultures were serially diluted on the corresponding agar plates for isolation. Sixteen colonies from each condition were chosen and used to inoculate tubes of corresponding media containing (trimethyl-d_9)-choline. The cultures were incubated for approximately 20 hours at 37 °C, and their TMA-d_9 levels were analyzed by LC-MS at the end of the incubation. 45 of 64 cultures produced significant amounts of TMA-d_9.

To further verify that the TMA produced by these cultures originated from choline, each of the isolates was grown on anoxic BHI media supplemented (or "P" media for isolate Pep7) with choline-d_9,

and TMA-d_9 was quantified as described using an alternate method that does not involve derivatization. Three 200 µL replicates were inoculated with 2 µL of a starter culture and incubated for 20 h at 37 °C under anaerobic conditions in a 96-well plate sealed with aluminum film. An aliquot of 50 µL of each culture was diluted 333 fold with 80% acetonitrile/ 20% 5 mM ammonium formate with 0.1% formic acid, and the concentration of TMA-d_9 in the culture media was calculated based on an external standard curve analyzed at concentrations ranging from 104 pg/µL to 627 pg/µL TMA-d_9. Isolates able to convert choline into TMA were chosen for 16S rRNA and *cutC* gene sequencing.

2.3.13: 16S rRNA sequencing

PCR amplification of 16S rRNA genes was performed on genomic DNA isolated from all of the new choline utilizing strains using previously reported specific primers, fD1 (5'-TCCTACGGGAGGCAGCAGT-3') and rP2 (5'- GGACTACCAGGGTATCTAATCCTGTT-3') (5). PCR reactions contained 1X PCR buffer, 0.2 mM dNTPs, 100 ng of DNA template, 100 pmoles of each primer, and 1.25 U of Takara Taq Hot Start Version (Takara Bio Inc.) in a total volume of 50 µL. Thermocycling was carried out in a MyCycler gradient cycler (Bio-Rad) using the following parameters: denaturation for 5 min at 95 °C, followed by 30 cycles of 10 s at 98 °C, 30 s at 50 °C, 2 min at 72 °C, and a final extension time of 5 min at 70 °C. PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide staining. PCR products were gel purified and sequenced to determine the identity and purity of the isolates. DNA sequences were analyzed with a BLASTN search against the Nucleotide collection (nr/nt). 16S rRNA gene sequences with >98% identity were classified as the same species.

2.4: References

1. Zeisel SH & da Costa KA (2009) Choline: an essential nutrient for public health. *Nutr Rev* 67(11):615-623.

2. Chen Y, Patel NA, Crombie A, Scrivens JH, & Murrell JC (2011) Bacterial flavin-containing monooxygenase is trimethylamine monooxygenase. *Proc Natl Acad Sci USA* 108(43):17791-17796.

3. Hippe H, Caspari D, Fiebig K, & Gottschalk G (1979) Utilization of trimethylamine and other *N*-methyl compounds for growth and methane formation by *Methanosarcina barkeri*. *Proc Natl Acad Sci USA* 76(1):494-498.

4. Seibel BA & Walsh PJ (2002) Trimethylamine oxide accumulation in marine animals: relationship to acylglycerol storage. *J Exp Biol* 205(3):297-306.

5. Chao CK & Zeisel SH (1990) Formation of trimethylamine from dietary choline by *Streptococcus sanguis* I, which colonizes the mouth. *J Nutr Biochem* 1(2):89-97.

6. Garsin DA (2010) Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nat Rev Microbiol* 8(4):290-295.

7. Thiennimitr P*, et al.* (2011) Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. *Proc Natl Acad Sci USA* 108(42):17480-17485.

8. Shibata N*, et al.* (2010) Crystal structures of ethanolamine ammonia-lyase complexed with coenzyme B_{12} analogs and substrates. *J Biol Chem* 285(34):26484-26493.

9. Yeates TO, Crowley CS, & Tanaka S (2010) Bacterial microcompartment organelles: protein shell structure and evolution. *Annual Review of Biophysics, Vol 39* 39:185-205.

10. Kofoid E, Rappleye C, Stojiljkovic I, & Roth J (1999) The 17-gene ethanolamine (*eut*) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. *J Bacteriol* 181(17):5317-5329.

11. Lehtio L & Goldman A (2004) The pyruvate formate lyase family: sequences, structures and activation. *Protein Eng Des Sel* 17(6):545-552.

12. Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696-704.

13. Selmer T, Pierik AJ, & Heider J (2005) New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem* 386(10):981-988.

14. Weimer PJ, Vankavelaar MJ, Michel CB, & Ng TK (1988) Effect of phosphate on the corrosion of carbon-steel and on the composition of corrosion products in two-stage continuous cultures of *Desulfovibrio desulfuricans*. *Appl Environ Microb* 54(2):386-396.

15. Luo QW, Groh JL, Ballard JD, & Krumholz LR (2007) Identification of genes that confer sediment fitness to *Desulfovibrio desulfuricans* G20. *Appl Environ Microb* 73(19):6305-6312.

16. Vey JL*, et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.

17. Unkrig V, Neugebauer FA, & Knappe J (1989) The free radical of pyruvate formate-lyase. Characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analog hypophosphite. *Eur J Biochem* 184(3):723-728.

18. Rabus R*, et al.* (2001) Anaerobic initial reaction of n-alkanes in a denitrifying bacterium: evidence for (1-methylpentyl)succinate as initial product and for involvement of an organic radical in n-hexane metabolism. *J Bacteriol* 183(5):1707-1715.

19. Pereira PM*, et al.* (2008) Energy metabolism in *Desulfovibrio vulgaris* Hildenborough: insights from transcriptome analysis. *Anton Leeuw Int J G* 93(4):347-362.

20. Nicholson JK*, et al.* (2012) Host-gut microbiota metabolic interactions. *Science* 336(6086):1262-1267.

21. Bain MA, Fornasini G, & Evans AM (2005) Trimethylamine: metabolic, pharmacokinetic and safety aspects. *Curr Drug Metab* 6(3):227-240.

22. Su C, Lei LP, Duan YQ, Zhang KQ, & Yang JK (2012) Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Appl Microbiol Biot* 93(3):993-1003.

23. Dobrindt U*, et al.* (2002) Genetic structure and distribution of four pathogenicity islands (PAI I-536 to PAI IV536) of uropathogenic *Escherichia coli* strain 536. *Infect Immun* 70(11):6365-6372.

24. Smillie CS*, et al.* (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480(7376):241-244.

25. Sommer MOA, Dantas G, & Church GM (2009) Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* 325(5944):1128-1131.

26. Louis P, McCrae SI, Charrier C, & Flint HJ (2007) Organization of butyrate synthetic genes in human colonic bacteria: phylogenetic conservation and horizontal gene transfer. *FEMS Microbiol Lett* 269(2):240-247.

27. Hehemann JH*, et al.* (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* 464(7290):908-U123.

28. Zeisel SH, Dacosta KA, Youssef M, & Hensey S (1989) Conversion of dietary choline to trimethylamine and dimethylamine in rats - dose-response relationship. *J Nutr* 119(5):800-804.

29. Brown SA, Palmer KL, & Whiteley M (2008) Revisiting the host as a growth medium. *Nat Rev Microbiol* 6(9):657-666.

30. Vital M, Howe AC, & Tiedje JM (2014) Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *MBio* 5(2): e00889.

31. Jones BV, Begley M, Hill C, Gahan CGM, & Marchesi JR (2008) Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc Natl Acad Sci USA* 105(36):13580-13585.

32. McIntosh FM*, et al.* (2012) Phylogenetic distribution of genes encoding beta-glucuronidase activity in human colonic bacteria and the impact of diet on faecal glycosidase activities. *Environ Microbiol* 14(8):1876-1887.

33. Toraya T (2014) Cobalamin-dependent dehydratases and a deaminase: radical catalysis and reactivating chaperones. *Arch Biochem Biophys* 544:40-57.

34. Wargo MJ (2013) Homeostasis and catabolism of choline and glycine betaine: lessons from Pseudomonas aeruginosa. *Appl Environ Microbiol* 79(7):2112-2120.

35. Muller E, Fahlbusch K, Walther R, & Gottschalk G (1981) Formation of N,N-dimethylglycine, acetic acid, and butyric acid from betaine by *Eubacterium limosum*. *Appl Environ Microb* 42(3):439-445.

36. Gaci N, Borrel G, Tottey W, O'Toole PW, & Brugere JF (2014) Archaea and the human gut: new beginning of an old story. *World J Gastroentero* 20(43):16062-16078.

37. Stoll S & Schweiger A (2006) EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J Magn Reson* 178(1):42-55.

38. Larkin MA*, et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.

39. Soding J, Biegert A, & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244-W248.

40. O'Brien JR*, et al.* (2004) Insight into the mechanism of the B_{12}-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* 43(16):4635-4645.

41. Tamura K*, et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731-2739.

42. Letunic I & Bork P (2011) Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* 39:W475-W478.

43. Wall JD, Murnan T, Argyle J, English RS, & RappGiles BJ (1996) Transposon mutagenesis in *Desulfovibrio desulfuricans*: development of a random mutagenesis tool from Tn7. *Appl Environ Microb* 62(10):3762-3767.

44. Krumholz LR & Bryant MP (1986) *Eubacterium oxidoreducens* sp. nov. requiring hydrogen or formate to degrade gallate, pyrogallol, phloroglucinol and quercetin. *Arch Microbiol* 144(1):8-14.

45. Johnson DW (2008) A flow injection electrospray ionization tandem mass spectrometric method for the simultaneous measurement of trimethylamine and trimethylamine N-oxide in urine. *J Mass Spectrom* 43(4):495-499.

46. Lamark T*, et al.* (1991) DNA sequence and analysis of the *bet* genes encoding the osmoregulatory choline glycine betaine pathway of *Escherichia coli*. *Mol Microbiol* 5(5):1049-1064.

Chapter 3: Characterization of Choline Trimethylamine-Lyase Expands the Chemistry of Glycyl Radical Enzymes[a]

3.1: Introduction

Nature harnesses radical chemistry to accomplish many challenging biochemical transformations (1). As described in Chapter 1, glycyl radical enzymes (GREs) utilize protein-based radical intermediates to catalyze a variety of reactions, including nucleotide reduction (class III ribonucleotide reductase (RNR)), C–C bond formation (benzylsuccinate synthase (BSS)), C–C bond cleavage (pyruvate formate-lyase (PFL) and 4-hydroxyphenylacetate decarboxylase (4-HPAD)), and dehydration (B_{12}-independent glycerol dehydratase (GDH)) (1). In this Chapter, we report the in vitro characterization of a new GRE from anaerobic microbes, the C–N bond cleaving enzyme choline trimethylamine (TMA)-lyase (CutC) (2). This work reveals the exquisite selectivity of CutC for choline cleavage and provides insights into the molecular basis for its function that will inform efforts to develop small molecule inhibitors of this microbial metabolic activity.

As detailed in Chapter 1, microorganisms in the human gastrointestinal tract and other anaerobic environments utilize choline as a source of carbon and energy. This process involves generation of TMA, an exclusively microbial metabolite (**Figure 3.1.A**) (3, 4). While choline is an essential nutrient for humans (5), excess conversion of choline to TMA by the gut microbiota and its further oxidation to trimethylamine *N*-oxide by liver enzymes is implicated in a number of diseases, including non-alcoholic fatty liver disease (6), atherosclerosis (7), and the metabolic disorder trimethylaminuria (fish malodor syndrome) (8). TMA is also a carbon source for bacteria and is converted by archaea to the greenhouse gas methane in various marine habitats (9).

[a] This chapter is an unofficial adaptation of the article: Craciun S, Marks JA, & Balskus EP (2014) Characterization of choline trimethylamine-lyase expands the chemistry of glycyl radical enzymes. *ACS Chem Biol* 9(7):1408-1413.

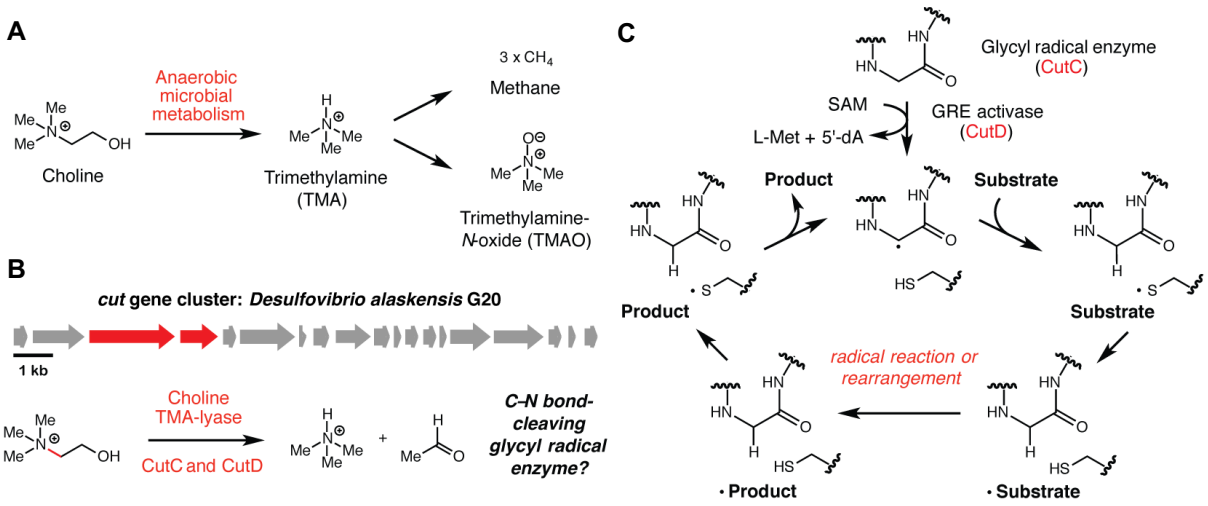


Figure 3.1: Choline trimethylamine-lyase (CutC): a predicted glycyl radical enzyme (GRE) involved in anaerobic choline metabolism. **(A)** Microbial generation of TMA from choline and its subsequent processing in biological systems. **(B)** The choline utilization (*cut*) gene cluster and the biochemical role of predicted GRE CutC and GRE activase CutD. **(C)** Shared mechanistic hypothesis for GRE function and proposed activities for CutC and CutD. SAM = S-adenosylmethionine; dA = 5'-deoxyadenosine; Met = L-methionine.

In Chapter 2, we presented the discovery of the first genes responsible for anaerobic microbial choline metabolism (2). The choline utilization (*cut)* gene cluster, which is widely distributed among sequenced bacteria and found in many human gut isolates, encodes a predicted bacterial microcompartment-associated GRE (CutC) and an S-adenosylmethionine (SAM) dependent GRE activating protein (CutD). We confirmed that CutC and CutD were responsible for the initial C–N bond fragmentation that generates TMA and acetaldehyde (**Figure 3.1.B**) using genetics and heterologous expression (2). As this transformation impacts human disease and represents a new reaction for the GRE family, we decided to further examine the activity of CutC and CutD in vitro.

The GREs characterized to date are hypothesized to share certain mechanistic features (**Figure 3.1.C**), as described in detail in Chapter 1 (1). The glycine-centered radical required for enzyme function is generated by the action of a separate activating protein, a member of the radical SAM enzyme family (10). These [4Fe-4S] cluster-containing enzymes homolytically cleave SAM, generating a 5'-deoxyadenosyl radical that abstracts an α-hydrogen atom from an active site glycine of a partner GRE. Substrate binding may trigger reaction of the glycyl radical with a nearby cysteine residue, possibly generating a thiyl radical that may initiate the reaction through formation of a substrate-based radical.

Following further reaction or rearrangement, a product-centered radical may react with the active site cysteine, regenerating the thiyl radical. The catalytic glycine and cysteine residues are conserved in all known GREs, including choline TMA-lyase, underscoring their importance for function, but there is not yet direct evidence for the thiyl radical intermediate. With this knowledge, we hypothesized that the putative glycyl radical activating enzyme CutD would generate a glycyl radical on CutC, and that upon activation, CutC would convert choline to TMA and acetaldehyde. In addition to reconstituting the choline-cleaving activity of these enzymes in vitro, we also sought to investigate their substrate specificity and identify additional active site residues that may contribute to catalysis.

3.2: Results and discussion

3.2.1: Overexpression and purification of CutC from D. alaskensis *G20 and* Escherichia coli *MS 69-1*

We initiated our studies by cloning CutC and CutD from *Desulfovibrio alaskensis* G20 genomic DNA and individually overexpressing the proteins in *Escherichia coli* as N- and C-His_6-tagged fusions, respectively. CutC was expressed and purified aerobically, and the full-length enzyme was employed for initial activity assays and all site-directed mutagenesis experiments. While scaling up the original purification procedure for analysis by electron paramagnetic resonance spectroscopy (EPR), we observed aggregation and precipitation during dialysis. Most CutC homologs are clustered together with genes encoding putative microcompartment structural proteins, suggesting that these enzymes might be microcompartment-associated in vivo. We thus hypothesized that the instability of full length CutC from *D. alaskensis* G20 could be due to aggregation in the absence of the microcompartment. Previous work with proteins from the 1,2-propanediol (11) and ethanolamine (12) utilization pathways revealed that short N-terminal sequences are responsible for packaging proteins into bacterial microcompartments. Removal of this 27 amino acid sequence led to an improved in vitro solubility of ethanolamine ammonia-lyase (EAL) subunit EutC from *E. coli*, without negatively impacting activity (12).

We performed a multiple sequence alignment of CutC from *D. alaskensis* G20 with several CutC homologs, three of which were not microcompartment-associated based on genomic context (**Figure 3.2**).

We observed that the N-terminal regions of these proteins have a lower degree of similarity (32%) than the rest of the sequences (74%). Thus, we proposed that the aggregation of full length CutC in vitro could be due to a short (up to ~50 amino acid) N-terminal sequence. Upon generation of truncated variants of CutC lacking the first 18, 30 or 52 amino acids (aa) (CutC–18 aa, CutC–30 aa, CutC–52 aa), we observed that these proteins possessed significantly improved solubility and stability in comparison to the full-length enzyme. As a result, the –52 aa truncated variant was used for EPR and kinetics experiments. Analytical gel filtration revealed a dimeric oligomerization state for full length CutC and all of the variants (**Figure 3.3**), a state similar to that of GDH (13), PFL (14), and class III RNR (15).

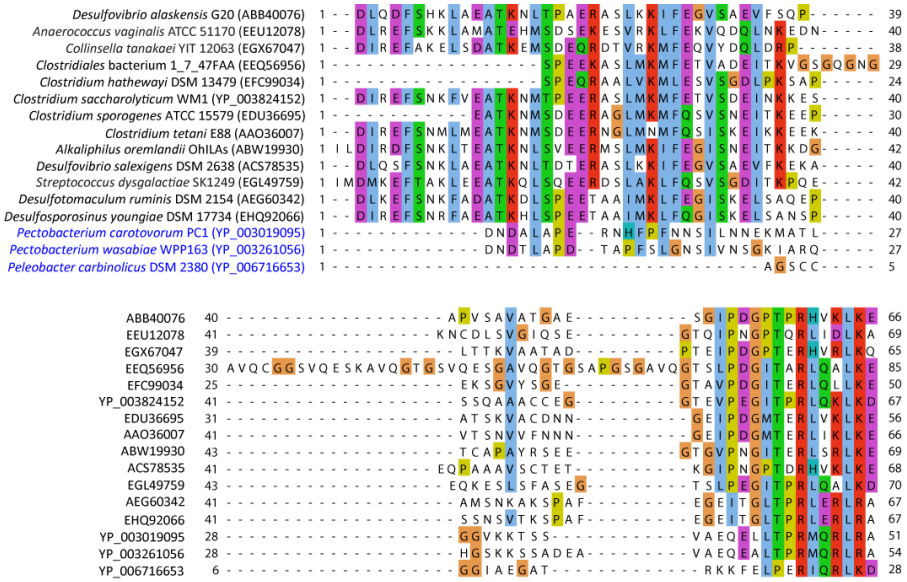


Figure 3.2: Multiple sequence alignment of the N-terminal regions of representative choline TMA-lyase homologs reveals that this region might target CutC to a protein microcompartment. The last three homologs (marked with blue) are part of *cut* operons that do not contain putative microcompartment proteins. N-terminal methionine residues were omitted to prevent spurious alignment of the divergent termini. The percent similarity of these regions is 32%, as opposed to 74% for the rest of the sequences.

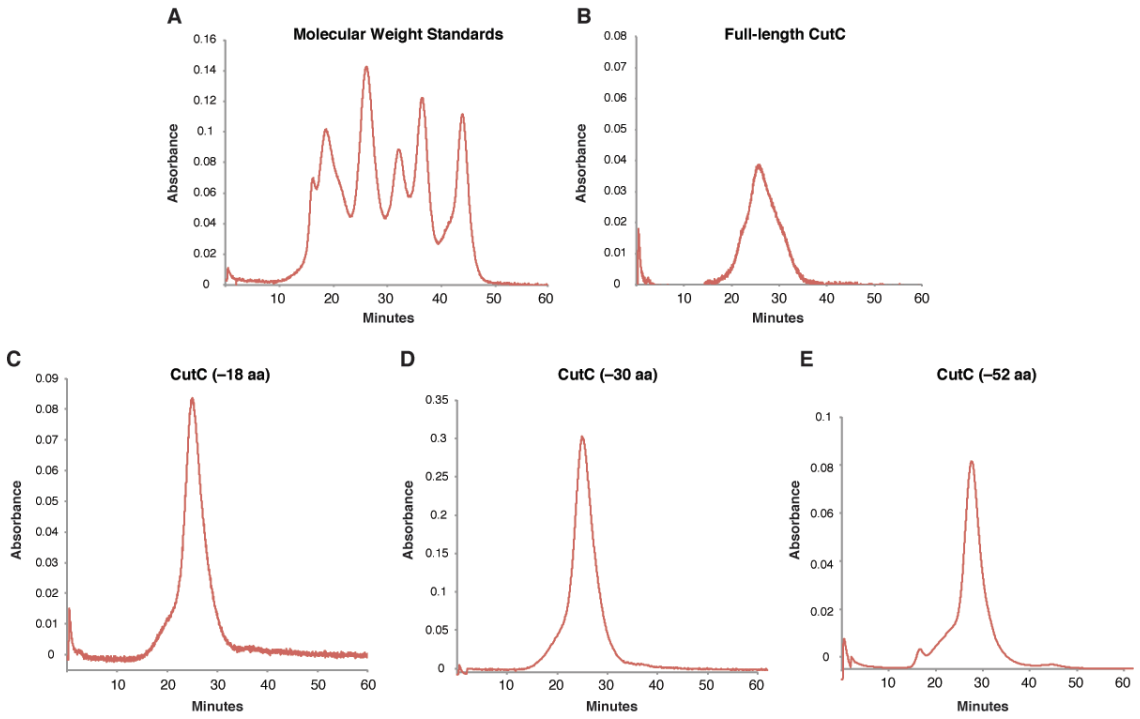


Figure 3.3: Determination of the oligomeric state of CutC from *Desulfovibrio alaskensis* G20 via size exclusion chromatography. **(A)** FPLC chromatogram of molecular weight standards, **(B)** full-length CutC, **(C)** CutC (–18 aa), **(D)** CutC (–30 aa), **(E)** CutC (–52 aa).

When searching for homologs of CutC from *D. alaskensis* G20 (2), we discovered that all of the putative choline TMA-lyases from facultative anaerobic bacteria contain an extra ~ 300 amino acid N-terminal domain that does not appear to share similarity with the rest of the protein. As described in Chapter 1, the GRE pyruvate formate-lyase from *E. coli* is encoded next to a small protein (YfiD) similar in sequence to the glycyl radical domain of PFL and which was found to rescue the activity of oxygen-damaged PFL. We cannot infer the function of the extra 300 amino acid N-terminal domain of CutC solely from function, so we aimed to biochemically characterize the full length and a N-terminal domain –334 aa truncation variant of CutC from *Escherichia coli* MS 69-1. Both proteins were purified to homogeneity as C-His_6-tagged fusions in very good yields. However, analytical size exclusion chromatography (**Figure 3.4**) revealed that the oligomeric state of these proteins was highly heterogeneous, with the predominant state as the tetrameric one.

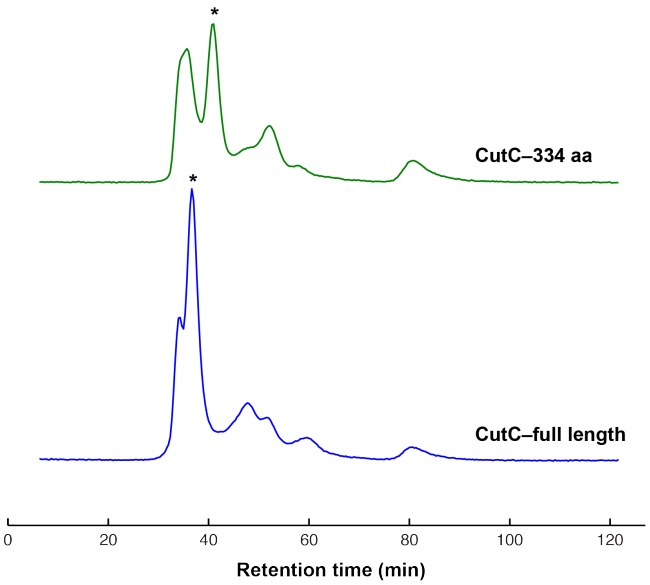


Figure 3.4: Determination of the oligomeric state of full length CutC and CutC–334 aa truncated variant from *Escherichia coli* MS 69-1 using analytical size exclusion chromatography. The major peak, marked with an asterisk, was found to correspond to the tetramer.

3.2.2: CutD is a SAM-dependent glycyl radical enzyme activating protein

Expression, purification, and reconstitution of putative GRE activase CutD from *D. alaskensis* G20 with a C-terminal His_6-tag were performed under strictly anaerobic conditions with co-expression of the *E. coli isc* operon (16). Interestingly, the co-expression of the *isc* operon improved CutD yields by ~2-fold, while the co-expression of the *suf* operon (typically used to repair oxidatively-damaged [4Fe-4S] cluster proteins) (17) interfered with the expression of CutD, providing greatly reduced yields of protein. Maltose binding protein-tagged and Strep-tagged CutD also purified in ~10-fold lower yields, so the C-His_6-tag construct was chosen for characterization. Analytical gel filtration revealed CutD to be a monomer, similarly to PFL-AE (10) (**Figure 3.5**). Multiple sequence alignment of CutD with characterized GRE activases revealed the conserved [4Fe-4S]-binding CX_3CX_2C motif, as well as putative SAM and GRE binding sites (**Figure 3.6**) (10). CutD also appeared to possess two "Clostridial"-type $CX_{2-5}CX_{2-4}CX_3C$ motifs, indicating the possibility of auxiliary [4Fe-4S] clusters (18). These motifs are typically present in ferredoxins, ubiquitous proteins in bacteria and archaea that typically mediate electron transfer within proteins (19). In the case of CutD, it is likely that the additional [4Fe-4S] clusters

could shuttle electrons from an electron donor (such as a reduced flavodoxin protein) to the SAM-bound [4Fe-4S] cluster, but generally, the roles of auxiliary [4Fe-4S] clusters in radical SAM enzymes are still poorly understood and difficult to study (20).

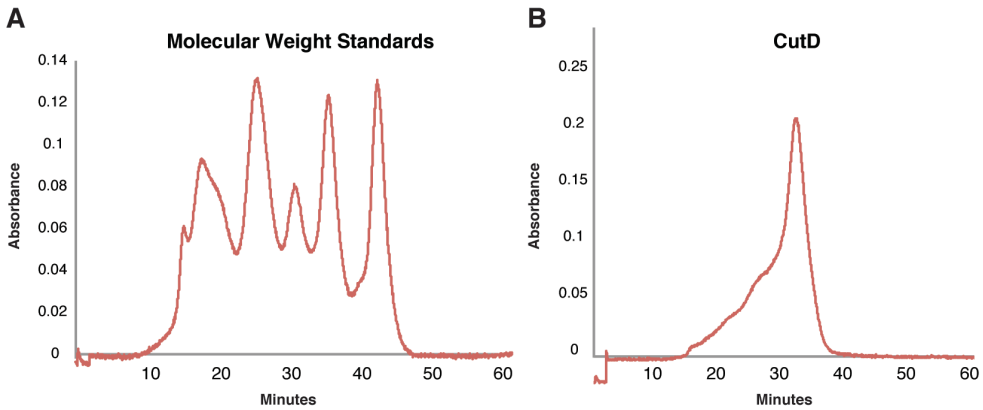


Figure 3.5: Determination of CutD molecular mass via size exclusion chromatography. **(A)** FPLC chromatogram of molecular weight standards. **(B)** FPLC chromatogram of CutD.

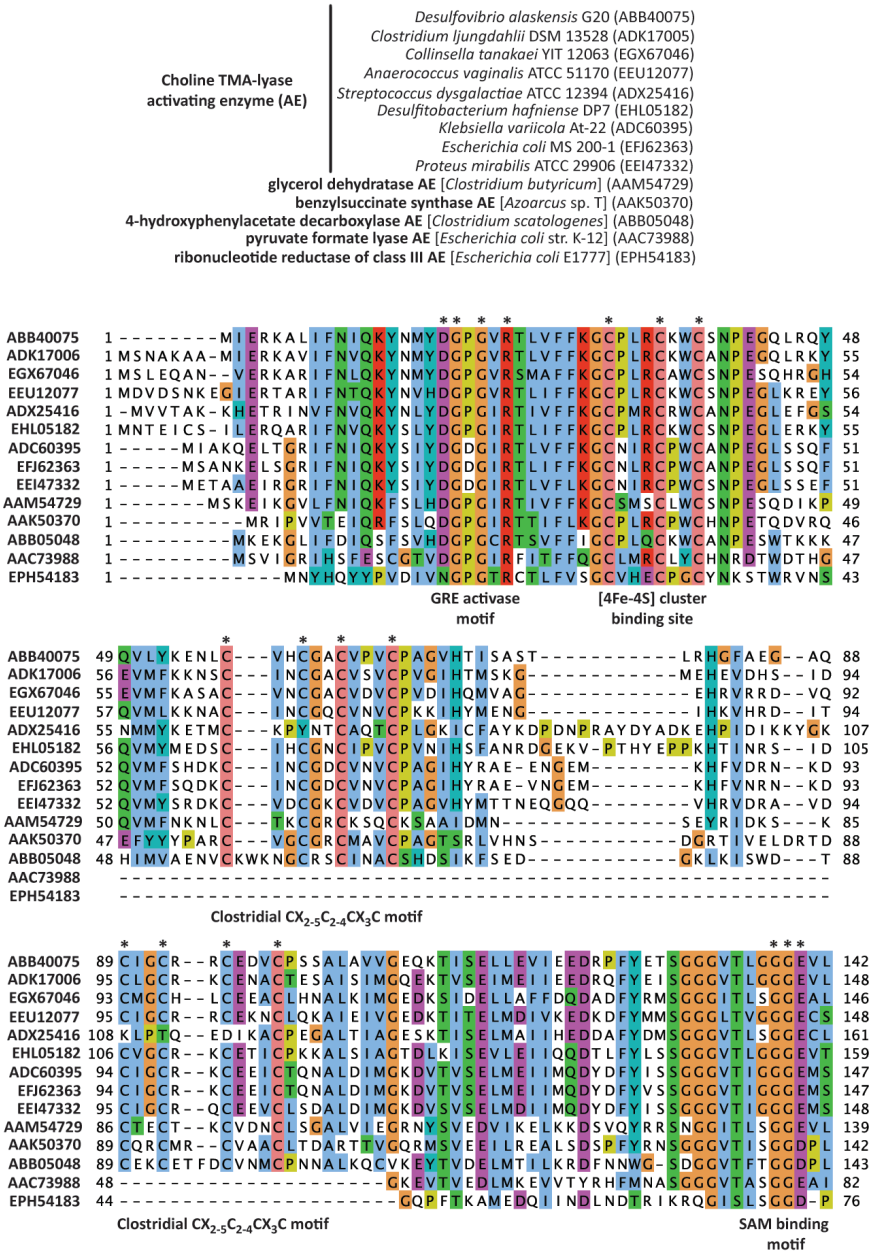


Figure 3.6: Multiple sequence alignment of putative choline TMA-lyase activating enzymes and known glycyl radical activating enzymes. Asterisks indicate conserved binding motifs. GenBank accession number identifies amino acid sequences used.

The UV-Vis spectrum of reconstituted CutD displayed features typical of $[4Fe\text{-}4S]^{2+}$ clusters: a broad peak at 380-420 nm which decreased in intensity upon reduction to the $[4Fe\text{-}4S]^{+}$ state using excess sodium dithionite (NaDT) (**Figure 3.7**) (21). Analyses of CutD-bound iron and sulfide indicated the possibility of as many as two [4Fe-4S] clusters (8.4 moles of Fe and 7.6 moles of S per mole of CutD), although we could not rule out adventitious binding of iron. The CutD extinction coefficient at 410 nm

prior to reduction was ~ 12.6 mM^{-1} cm^{-1}, suggesting that not all iron-sulfur centers had been properly reconstituted ($[4Fe–4S]^{2+}$ ε_{410} ~15 mM^{-1} cm^{-1}) (21). The EPR spectrum of NaDT-reduced CutD supported this hypothesis. At 9 K, the spectrum contained an axial signal with g|| of 2.045 and g⊥ of 1.94 which broadened at 40 K and did not change upon addition of excess SAM (**Figure 3.8**). The g-values and temperature dependence of this signal further confirmed the presence of $[4Fe\text{-}4S]^{+}$ centers in reduced CutD, while its intensity (0.8 spins per CutD monomer) is consistent with incomplete reconstitution. We also examined the activity of NaDT-reduced CutD toward SAM in the absence of CutC using high performance liquid chromatography (HPLC) assays (22) and observed generation of cleavage products 5'-deoxyadenosine and L-methionine, as previously found for other GRE activating proteins (23) (**Figure 3.9**). Together, these analyses confirm the assignment of CutD as a radical SAM enzyme.

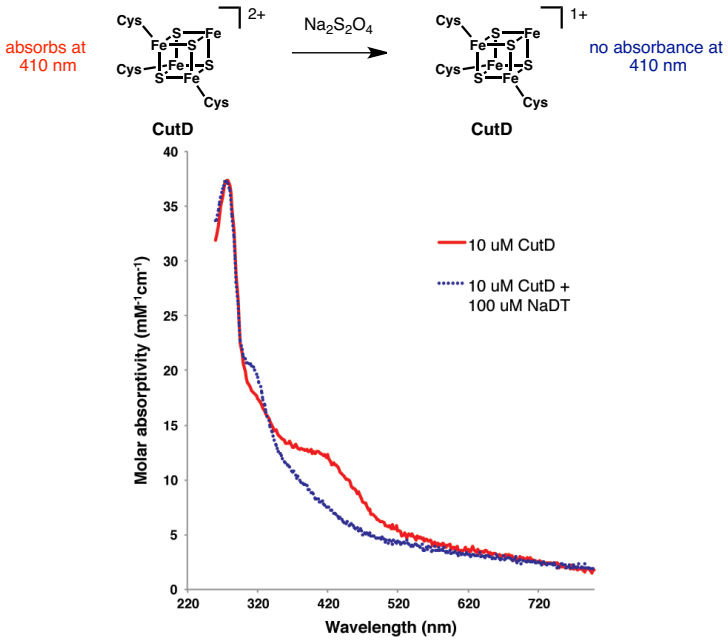


Figure 3.7: UV-Vis spectrum of purified CutD before (red) and after (blue) reduction with sodium dithionite (NaDT).

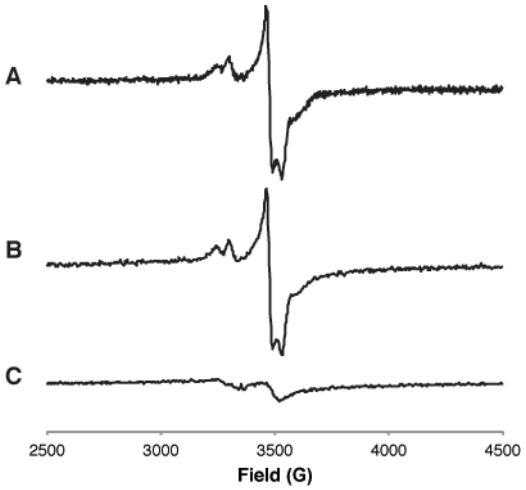


Figure 3.8: EPR characterization of the iron-sulfur cluster in CutD. **(A)** Spectrum at 9K of NaDT-reduced CutD. **(B)** Spectrum at 9 K of NaDT-reduced CutD incubated for < 5 min with excess SAM. **(C)** Spectrum at 40 K of NaDT-reduced CutD. Acquisition parameters are described in 'materials and general methods'.

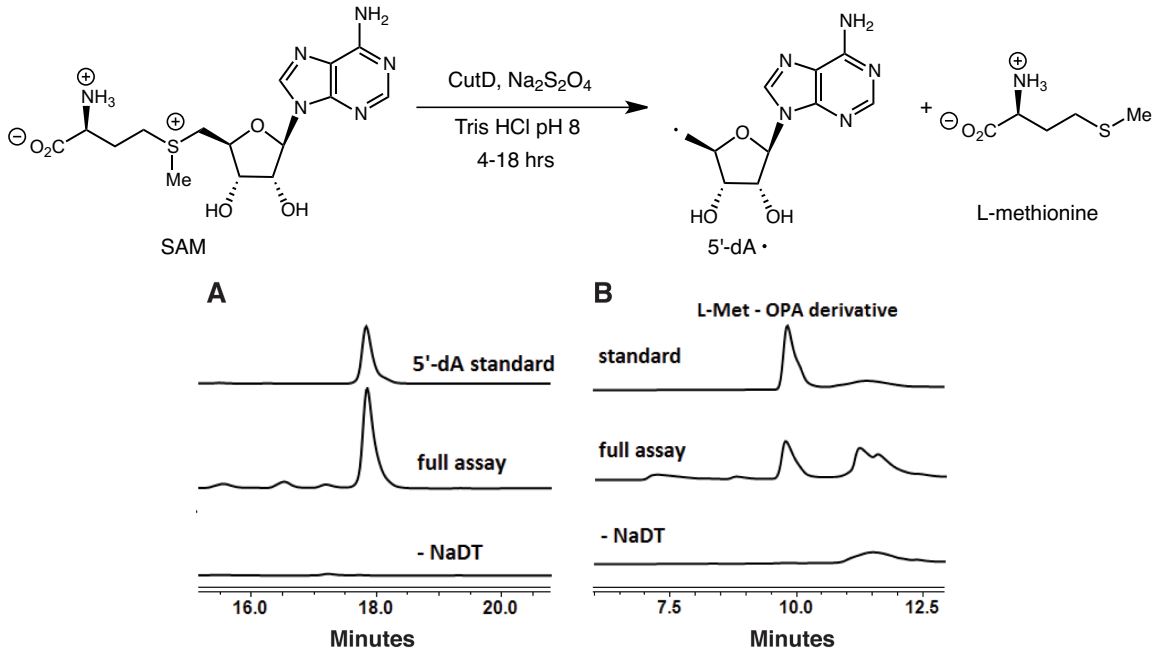


Figure 3.9: HPLC assays for characterization of SAM cleavage products generated in the presence of CutD. **(A)** Detection of 5'-deoxyadenosine (5'-dA), after the 5'-dA• is quenched by the solvent. **(B)** Detection of L-methionine as an OPA derivative.

3.2.3: The α-proton of the CutC glycyl radical is exchangeable with solvent

We performed additional experiments to investigate the reactivity of CutD toward CutC. Spectra of anaerobic reaction mixtures containing CutD, CutC, SAM, and NaDT taken at 77 K revealed a signal centered at g = 2.0037 and characterized by a 2-fold splitting of ~1.46 mT (**Figure 3.10.A**). The line

shape of this signal and its g value closely resemble those of the glycine-centered radicals from activated PFL (24) and class III RNR (15). Under the conditions described in 'materials and general methods', we observed activation to the glycyl radical species of only 9.3% of the CutC–52 aa truncated polypeptides from *D. alaskensis* G20, and 3.6% and 3.4% of CutC–full length and –334 aa truncated polypeptides from *E. coli* MS 69-1, respectively. Activation of CutC–52 aa from *D. alaskensis* G20 was not improved by employing catalytic amounts of CutD or by including choline in the reaction mixture. Use of the NADPH:flavodoxin reductase – flavodoxin system from *E. coli* (25) reduced glycyl radical formation (<1 % of the CutC dimer), and 5-deazariboflavin did not prove to be a suitable reductant for CutD, contrary to the results obtained for the activation of PFL (24).

When the CutC activation assay was run in Tris-buffered D_2O, the resulting EPR spectrum revealed no hyperfine splitting, indicating that the remaining α-proton of the glycyl radical is exchangeable with solvent (**Figure 3.10.B**). Analogous behavior, which is likely unrelated to the reaction mechanism, has been observed for other GREs (PFL (24), BSS (26), and 4-HPAD (27)) and has been attributed to participation of the conserved cysteine via thiyl radical formation. A CutC mutant lacking this cysteine (C489A) was still activated by CutD (**Figure 3.10.C**) but did not undergo exchange in D_2O (**Figure 3.10.D**), consistent with this residue's involvement in the exchange. These EPR experiments therefore not only confirm the biochemical function of CutD as a GRE activase, but also support the hypothesized functional similarities between CutC and other members of the GRE family.

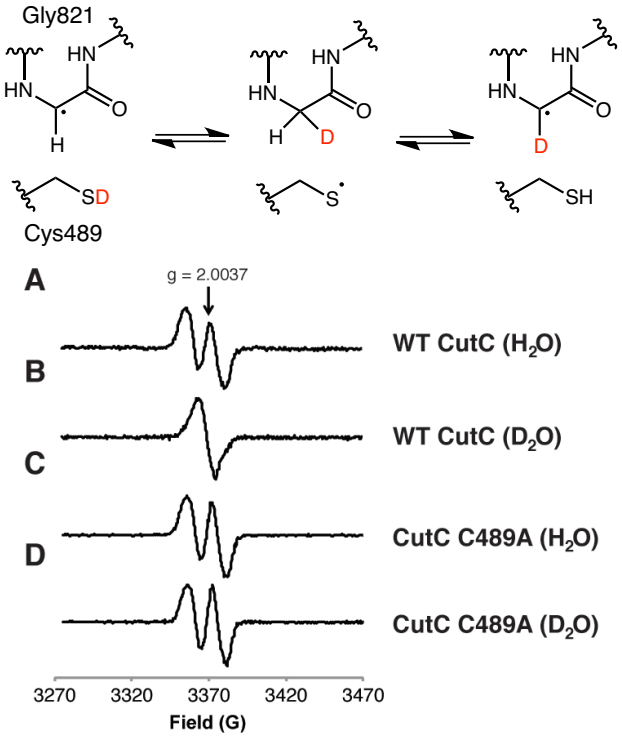


Figure 3.10: EPR spectra of activated CutC from *D. alaskensis* G20 reveals a glycyl radical with an α-proton that is exchangeable in the presence of Cys489. Experimental EPR signals of wild type CutC (–52 aa) activated in **(A)** buffered H_2O (g = 2.0037, A = 1.46 mT) and **(B)** buffered D_2O (g = 2.0035, A = 0.18 mT). Experimental EPR signals of C489A mutant CutC (–52 aa) activated in **(C)** buffered H_2O (g = 2.0033, A = 1.48 mT), and **(D)** buffered D_2O (g = 2.0033, A = 1.46 mT). The indicated g-values and isotropic hyperfine coupling (A-value) were deduced from simulations of the experimental spectra. Acquisition parameters and simulations are detailed in 'materials and general methods'.

3.2.4: CutC is a C–N cleaving glycyl radical enzyme

We initially examined the reactivity of the two-enzyme system using an end point assay containing choline, CutC and CutD from *D. alaskensis* G20, SAM, and NaDT. We observed complete conversion of choline to TMA (via LC-MS/MS) and acetaldehyde (via GC-MS/MS) only in strictly anaerobic assays that contained all of the reaction components, irrespective of the truncated CutC variant used (**Figures 3.11, 3.12, Table 3.1**). Together with EPR data, these experiments confirm that CutC is a C–N bond cleaving GRE. In order to assess the catalytic parameters of CutC-catalyzed choline cleavage and determine differences in activity of CutC from obligate anaerobic vs. facultative anaerobic bacteria, we performed a continuous spectrophotometric coupled assay, in which the formation of acetaldehyde from choline by activated CutC was coupled to the reduction of acetaldehyde by NADH-dependent yeast alcohol dehydrogenase (YADH) (13). We observed consistent activity for the CutC –52 aa truncated

variant from *D. alaskensis* G20, and CutC–full length and –334 aa truncated variant from *E. coli* MS 69-1. All proteins displayed kinetic parameters similar to each other and in accordance to those reported for other GREs (**Table 3.1, Figure 3.13**). It appears that under the experimental conditions we used, CutC from *E. coli* has a 5-fold greater affinity towards choline than CutC from *D. alaskensis*. The in vivo relevance of these values is unclear, as these experiments cannot replicate the environment of the bacterial microcompartment, which is likely to control substrate concentrations (2).

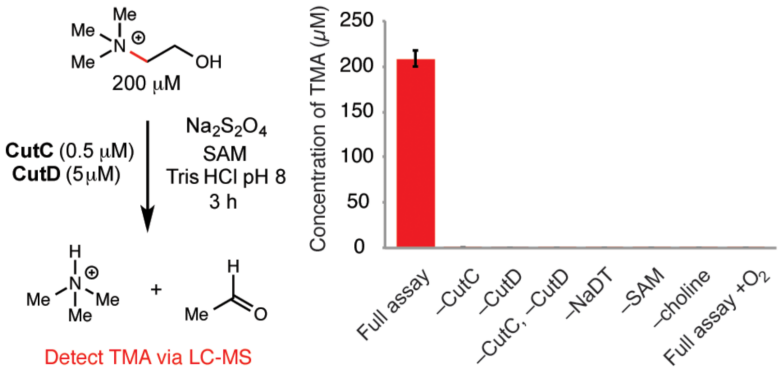


Figure 3.11: Reconstituted CutC from *D. alaskensis* G20 cleaves choline to TMA and acetaldehyde. LC-MS end point assays quantifying TMA production from choline by CutC. Each bar represents the mean ± standard error of the mean (SEM) of three replicates.

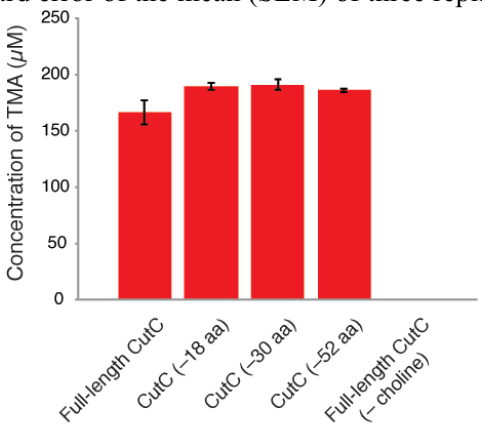


Figure 3.12: LC-MS assays detecting TMA produced from cleavage of choline by full length CutC (purified with either Tris-HCl or potassium phosphate buffers) and truncated CutC variants from *D. alaskensis* G20. Each bar represents the mean ± standard error of the mean (SEM) of three replicates.

Table 3.1: Kinetic parameters of CutC and comparison to those of other glycyl radical enzymes.

Enzyme	Substrate	K_m (μM)	V_{max} (μmol min^{-1} mg^{-1})	k_{cat} (s^{-1})[a]	k_{cat}/K_m (μM^{-1} s^{-1})
CutC–52 aa *D. alaskensis* G20	choline	302.5 ± 7.9	22.7 ± 0.2	374 ± 14	1.24
CutC–full length *E. coli* MS 69-1	choline	60 ± 0.01	16.1 ± 0.4	239 ± 6	3.98
CutC–334 aa *E. coli* MS 69-1	choline	63 ± 0.01	25.9 ± 0.6	282 ± 6	4.48
4-HPAD	HPA	649 ± 90	15.0 ± 0.2	110[b]	0.17[c]
PFL	pyruvate	2050	16.6	1100[b]	0.536[c]
BSS	toluene	< 100	0.0074	N/A	N/A
GDH	1,2-propanediol	N/A	1560	N/A	N/A
class III RNR	CTP (with 0.1 mM ATP as modulator)	450	1.4	6.8[c]	0.015[c]

[a] Turnover number shown is based on the amount of active protein. [b] Value previously reported.
[c] Value calculated from published data.

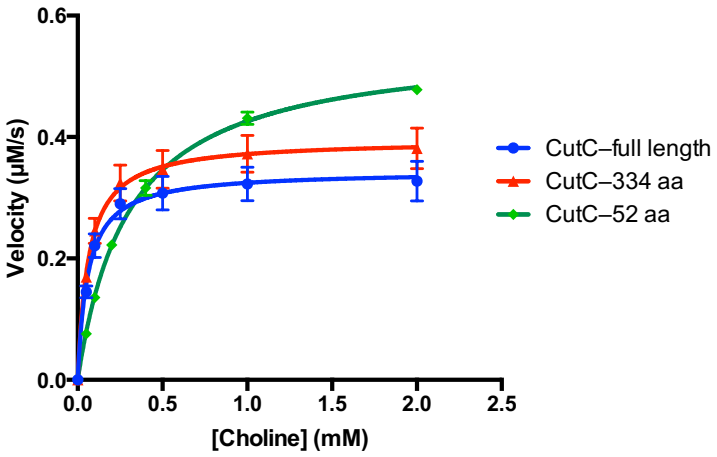


Figure 3.13: Michaelis-Menten kinetics of CutC–52 aa truncated variant from *D. alaskensis* G20, and CutC–full length and –334 aa truncated variant from *E. coli* MS 69-1.

3.2.5: Investigation of CutC specificity toward alternate substrates

Several CutC homologs had been identified prior to our work via bioinformatics and were postulated to mediate diol dehydration, potentially calling into question our annotation of CutC as a choline-cleaving GRE (28). We thus investigated the specificity of CutC by examining its activity toward alternate substrates using LC-MS/MS for TMA formation and GC-MS/MS for aldehyde formation. As

described below, we were unable to detect cleavage products in any assays involving physiological relevant small molecules such as L-carnitine, propanediol, dimethylaminoethanol, or glycine betaine (**Figures 3.14.A**, **3.15, 3.17.A, B**). The results of preliminary end point assays indicated that CutC can catalyze the C-N cleavage of both 2-methylcholine enantiomers (**Figure 3.16**) and *N*-(2-hydroxyethyl)-*N,N*-dimethylpropan-1-aminium (**Figure 3.17.D**), but we have not yet examined the kinetic parameters of these substrates. The high selectivity of CutC therefore disproves the hypothesis that these enzymes are diol-cleaving GREs. In addition, it appears that the presence of a trimethylammonium group is necessary, but not sufficient, for CutC-catayzed C–N cleavage of a small molecule. Overall, these results indicate that CutC is likely a specific molecular marker for anaerobic choline metabolism.

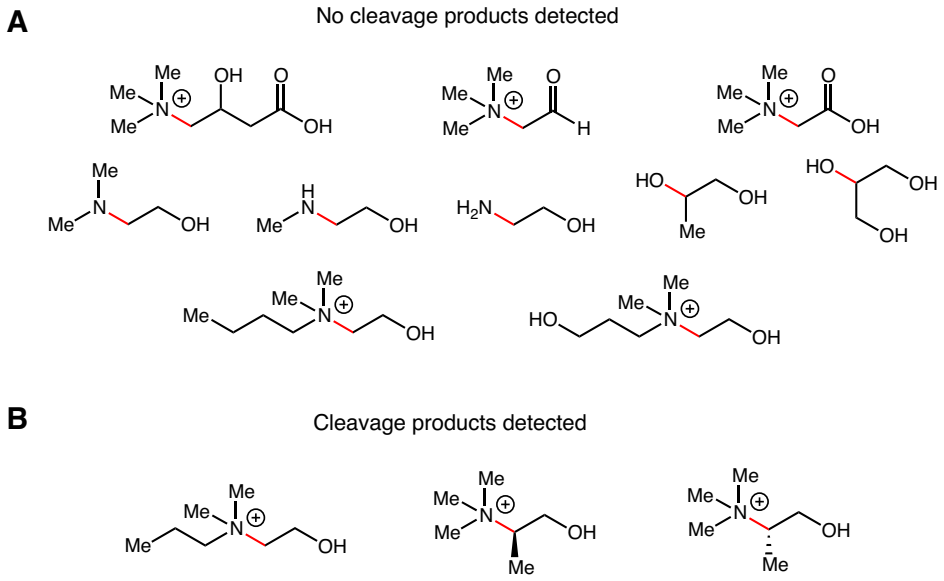


Figure 3.14: Substrates examined for cleavage by CutC using LC-MS/MS for trimethylamine analysis or GC-MS/MS for acetaldehyde, propionaldehyde, or 3-hydroxypropionaldehyde analyses. No cleavage products were detected for the molecules in **(A)**, whereas the molecules in **(B)** were cleaved by activated CutC.

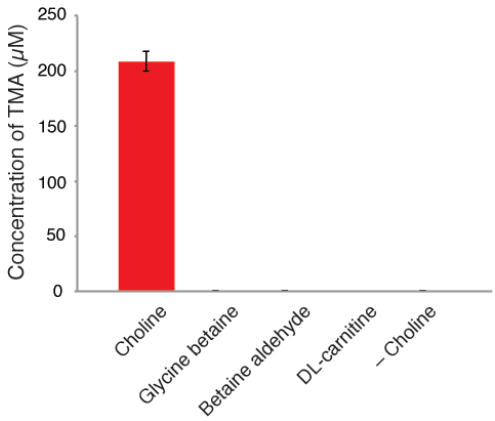


Figure 3.15: LC-MS assays detecting TMA produced from cleavage of choline, glycine betaine, betaine aldehyde, or DL-carnitine by CutC. Each bar represents the mean ± standard error of the mean (SEM) of three replicates.

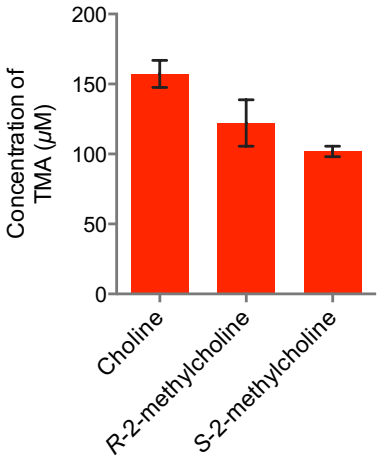


Figure 3.16: LC-MS assays detecting TMA produced from cleavage of choline, (*R*)-2-methylcholine and (*S*)-2-methylcholine by CutC. Each bar represents the mean ± standard deviation of two replicates.

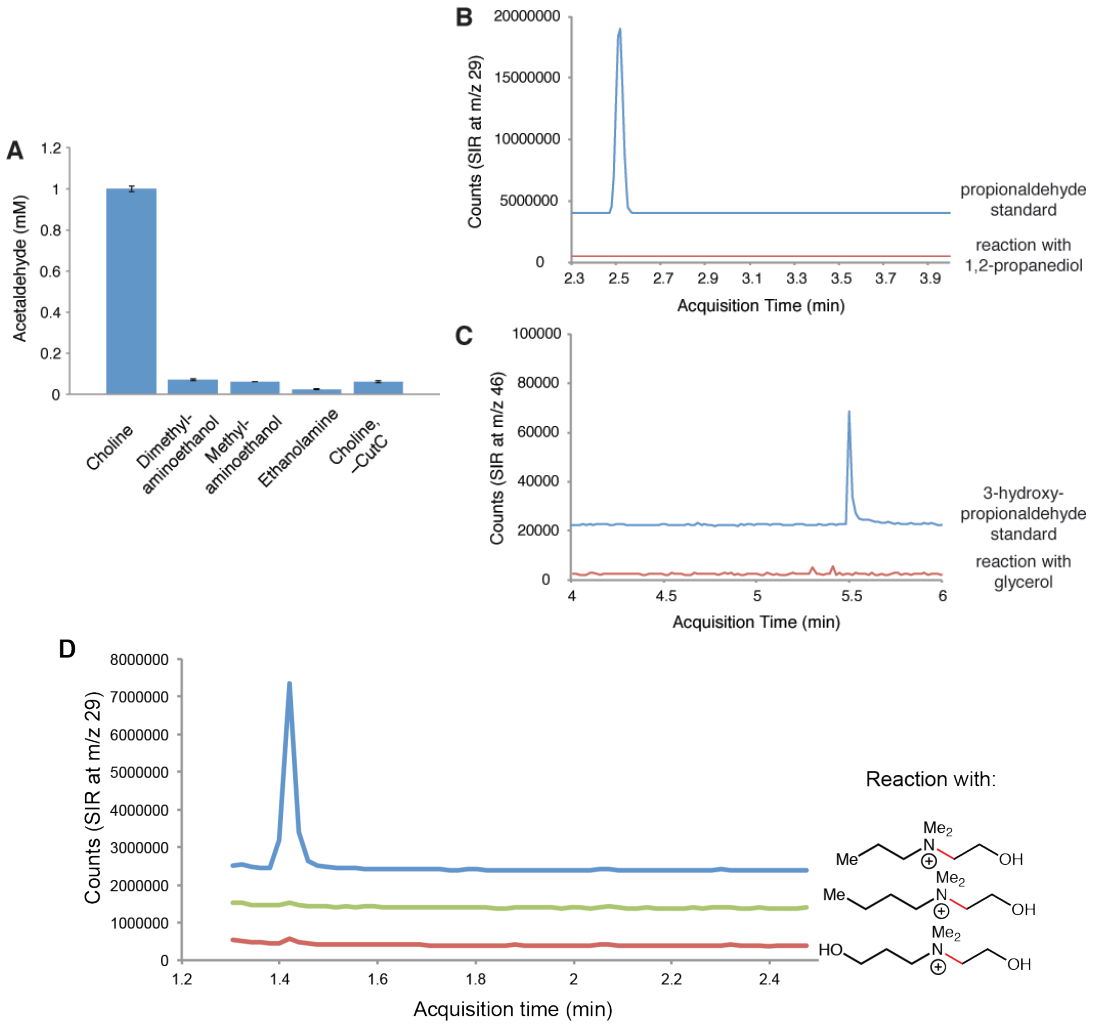


Figure 3.17: GC-MS assays for CutC activity with choline and other potential substrates. **(A)** Quantification of acetaldehyde production with alternate substrates. Bar graphs represent the mean ± standard error of the mean (SEM) of three independent experiments. **(B)** Selected ion recording (SIR) chromatogram for CutC incubated with 1,2-propanediol. **(C)** SIR chromatogram for CutC incubated with glycerol. **(D)** SIR chromatogram for CutC incubated with indicated alternate substrates.

3.2.6: Homology modeling of CutC structure and choline binding

To better understand the molecular basis for this specificity, we identified CutC active site residues involved in substrate binding and catalysis. An undergraduate student in the group, Jonathan Marks, built a series of CutC homology models using Modeller (29). The model chosen for further evaluation utilized two GDH structures as templates (PDB IDs: 1R8W and 2F3O) and passed multiple structural quality benchmarks (**Table 3.2, Figure 3.18**). He also utilized Schrödinger Suite 2012 (30) for active site refinement and induced-fit docking of choline (**Figure 3.19**). Overall, the predicted CutC active site closely resembled that of other GREs, with the conserved glycine and cysteine residues

(Gly821 and Cys489) located in identical positions. The orientation of choline in the active site revealed multiple contacts that could contribute to binding and catalysis. Some of these interactions, such as the hydrogen bonding between Glu491 and the choline C1 hydroxyl group, are also observed in the crystal structure of glycerol-bound GDH (13), suggesting that both enzymes may use Glu491 as a general base. However, the contacts between the protein scaffold and the trimethylammonium group appear unique to CutC and might help to distinguish it from other GREs: an electrostatic interaction with Asp216, hydrophobic packing with the methyl groups of Thr334 and Thr502, and a cation-π interaction with Phe395. Similar interactions were also observed in crystal structures of other choline and acetylcholine binding proteins (31, 32), suggesting that proteins might have evolved a common strategy to bind substrates containing a trimethylammonium group.

Table 3.2: Comparison of structure quality benchmarks for CutC homology model with those for the crystal structures of GDH (1R8W) and the predicted GDH GRE from *A. fulgidus* (2F3O).

Overall Scoring Metrics	**CutC Model**	**1R8W**	**2F3O**
DOPE (from Modeller)	-101582.21[a]	-104901.53	-107507.26
Verify3D average per residue	0.35417[a]	0.45087	0.45981
DFIRE2 Energy	-1557.36	-1584.73	-1613.4
dDFIRE Energy	-2019.05	-2046.52	-2079.01
QMean6	0.776	0.808	0.774
QMean6 z-score	0.142	0.494	0.125
ERRAT Quality Factor	78.777[a]	97.301	94.51
PROVE Score	0.988	1.055	1.414
Backbone conformation z-score	-4.518	-5.185	-3.377
Ramachandran			
Z-score (WHATCHECK)	1.46	-2.312	-0.969
% of residues in favorable regions	91.9	90.4	88.4
% of residues in allowed regions	8	9.5	11.6
% of residues in disallowed regions	0.1	0.1	0

[a]Denotes metrics that are particularly affected by the presence of the long disordered N-terminal tail of CutC.

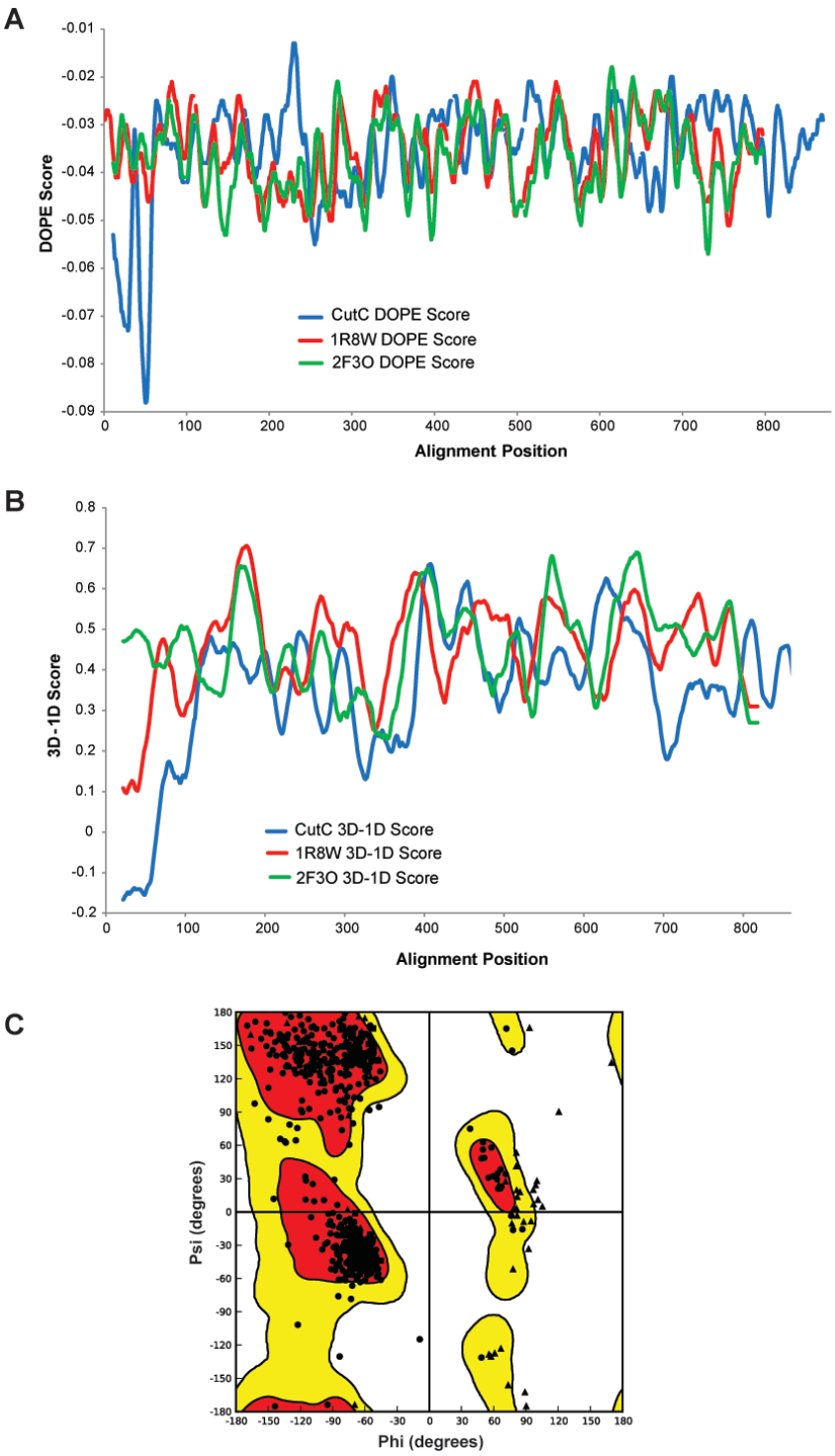


Figure 3.18: (A) Comparison of the DOPE profiles for the CutC homology model (blue), the published GDH structure (1R8W, red), and the predicted GDH (*A. fulgidus*) structure (2F3O, green). **(B)** 3D-1D profile of the CutC homology model (blue), the published GDH structure (red), and the predicted GDH (*A. fulgidus*) structure (green). **(C)** Ramachandran plot of homology modeled CutC structure, generated with Schrödinger Maestro. Most favored regions are shown in red. Additional allowed regions are shown in yellow. Disallowed regions are shown in white.

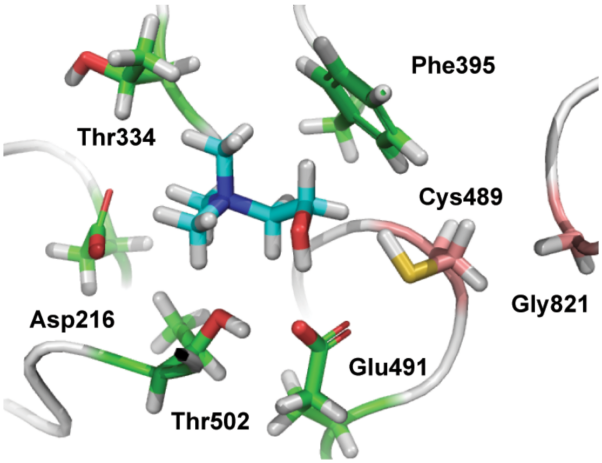


Figure 3.19: Docking of choline in the active site of a CutC homology model.

3.2.7: Preliminary investigation of the role of CutC active site residues in catalysis

We investigated the contributions to enzyme function of CutC active site residues identified in the homology model using site-directed mutagenesis of full length CutC. We aimed to change these residues to isosteric ones that lack the hypothesized functionality. Although all of the mutants expressed well, only T334S and T502S retained the ability to generate TMA from choline in an endpoint assay (**Figure 3.20**). At the time of this assay, we did not possess data regarding the extent of activation of each mutant to the glycyl radical form. Therefore, we assumed that incubating each mutant CutC with the activation assay components and choline for 3 hours would be sufficient to determine whether the mutation had an effect on its ability to cleave choline. We later realized that our procedure for activating these mutants to the glycyl radical form was suboptimal, and thus the lack of activity could have also stemmed from a defect in activation.

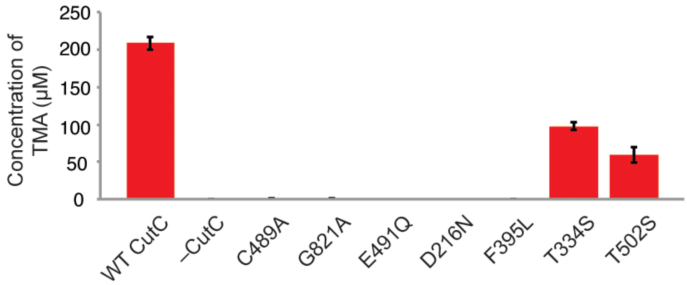


Figure 3.20: LC-MS end point assays quantifying TMA production from choline by by CutC active site mutants. Each bar represents the mean ± SEM of three replicates. WT = wild type.

Multiple sequence alignments of CutC homologs (2) (60-80% aa similarity) revealed conservation of all the active site residues identified in the homology model for CutC. Notably, alignments of CutC with other characterized GREs indicated that the combination of amino acids hypothesized to interact with the trimethylammonium group (Asp216, Thr334, Phe395, and Thr502) is not found in the active sites of other GREs **(Figure 3.21)**. The discovery of CutC amino acids that are both unique and required for function further supports the use of this sequence as specific marker for anaerobic choline metabolism.

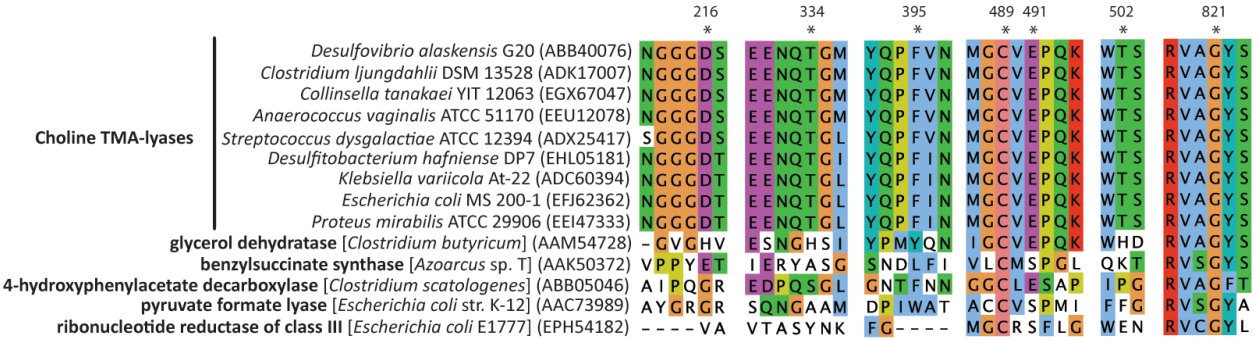


Figure 3.21: Multiple sequence alignment of choline TMA-lyase homologs and characterized glycyl radical enzymes. Asterisks indicate putative CutC active site residues. GenBank accession number identifies amino acid sequences used.

3.2.8: Mechanistic proposals for CutC-catalyzed choline cleavage

Combining the knowledge gained from modeling and mutagenesis experiments, and precedent from other radical-utilizing enzymes, we propose two preliminary mechanistic hypotheses for CutC-catalyzed choline cleavage (**Figure 3.22**). As is postulated for other GREs (1), the glycyl radical of activated CutC may react with Cys489 to form a thiyl radical. Subsequent hydrogen atom abstraction from C1 of choline would produce an initial substrate-based radical. Elimination of TMA, potentially involving a radical cation or radical anion intermediate, would generate a vinoxy radical that could abstract a hydrogen atom from Cys489 to generate acetaldehyde. This scenario closely resembles the proposed C–O bond-cleaving step in the mechanism of RNR (33); calculations have also suggested a similar mechanism would be possible for B_{12}-independent GDH (34). However, at this point, our data did not rule out an alternative scenario involving heteroatom migration that would be analogous to the generally accepted mechanisms of B_{12}-dependent eliminating enzymes, including ethanolamine ammonia-

lyase (35). We also do not yet have any direct evidence for participation of a thiyl radical in catalysis. As described in Chapter 4, we conducted a structural study of CutC, which will likely inform further mechanistic and computational efforts aimed at fully illuminating the molecular details of choline cleavage by CutC.

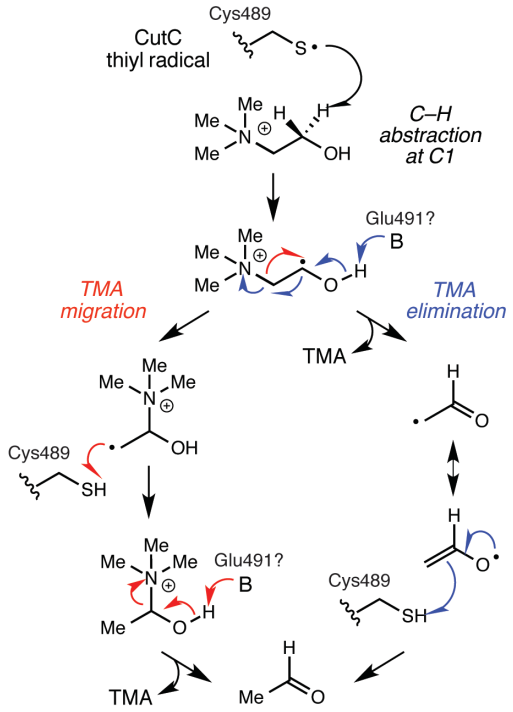


Figure 3.22: Proposed mechanisms for CutC-catalyzed choline cleavage.

3.2.9: Reactivity of CutC toward C1-^{13}C-labeled choline and stereospecifically-labeled [1-^{2}H]-choline

As an additional test of our docking model and an initial probe of mechanism, we examined the reactivity of CutC toward [1-^{13}C]-labeled choline, (*R*)-[1-^{2}H]-choline and (*S*)-[1-^{2}H]-choline. We were able to purchase [1-^{13}C]-choline and use GC-MS/MS to compare the ^{13}C enrichment of the acetaldehyde resulting from overnight incubation of activated CutC with either labeled or unlabeled choline. Based on the MS/MS spectra presented in **Figure 3.23**, we concluded that ^{13}C was fully incorporated into C1 of acetaldehyde. This result is consistent with an initial C-H abstraction occurring at this position to generate a C1-centered radical, in agreement with mechanisms proposed for B_{12}-independent glycerol dehydratase (34), and ethanolamine ammonia-lyase (35).

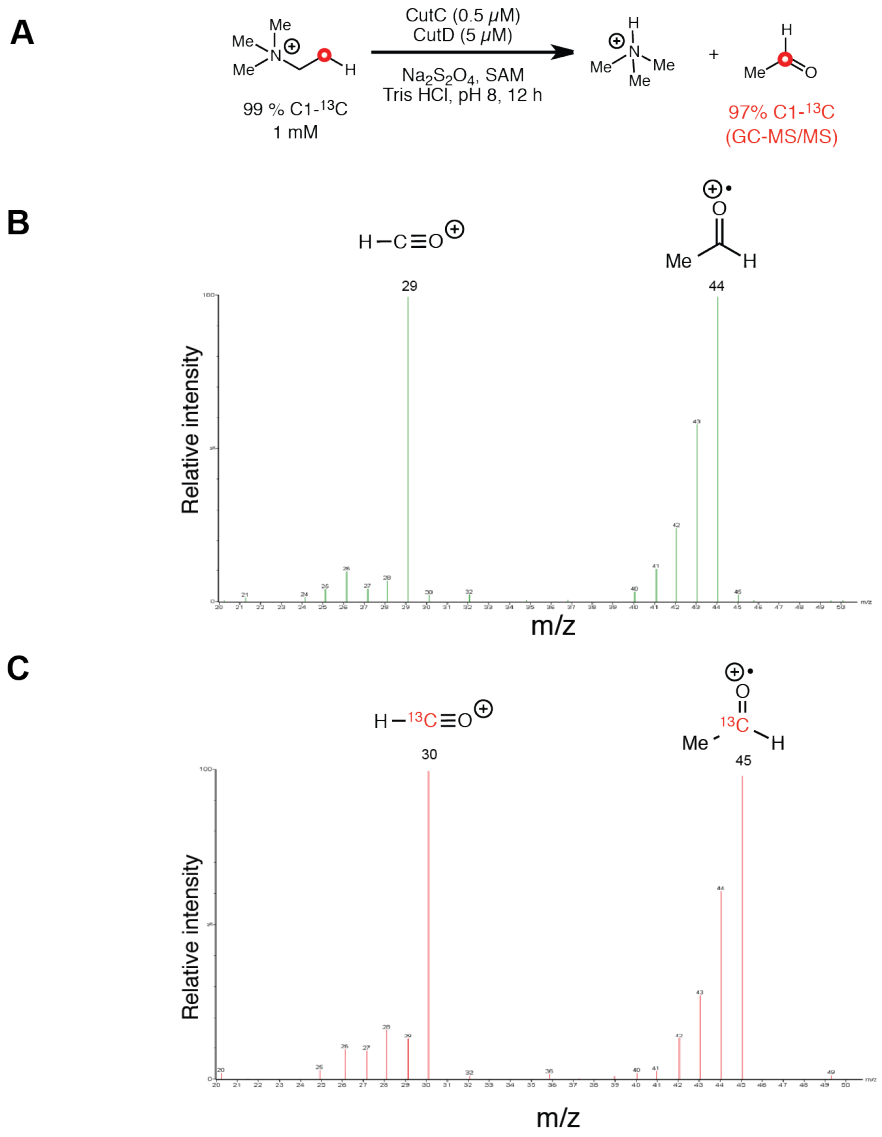


Figure 3.23: Reactivity of CutC toward ^{13}C-labeled choline. **(A)** General scheme for the assay. **(B)** MS/MS spectrum of acetaldehyde generated from reaction of CutC with unlabeled choline. **(C)** MS/MS spectrum of acetaldehyde generated from reaction of CutC with [1-^{13}C]-choline.

The orientation of choline within the active site of our homology model for CutC suggested that the putative thiyl radical on Cys489 might preferentially access the pro-*S* hydrogen atom at C1 for initial abstraction. By either proposed mechanism, the abstracted C1 hydrogen would be added to C2 after C–N cleavage, provided exchange does not occur with solvent or proton acceptors adjacent to Cys489 during the course of the reaction. In addition, the hydrogen atom not undergoing abstraction would remain on C1. Aiming to verify this hypothesis, we drew inspiration from a similar study of EAL-catalyzed cleavage

of deuterated ethanolamine analogs (36), and accessed deuterium labeled (*R*)-[1-^{2}H]-choline and (*S*)-[1-^{2}H]-choline, according to the synthetic route detailed in **Figure 3.24**. Our initial attempts to work with starting materials that had an unprotected amine group did not afford the desired products, so we decided to utilize the tert-butyloxycarbonyl (Boc) protecting group. Stereoinduction was achieved by reducing a deuterium labeled Boc-protected 2-aminoacetaldehyde with either (*R*) or (*S*)-Alpine Borane (37). Due to issues with aldehyde stability, we performed this reduction without prior purification. In addition, we observed that acid-catalyzed removal of the Boc group could lead to the formation of black, polymerized side products, which were difficult to purify away from the desired product. Future efforts to optimize this synthesis should thus focus on exploring alternate protecting strategies, such as a carboxybenzyl (Cbz) group. Nonetheless, we were able to obtain the desired deuterium labeled choline analogs, and we determined their purity and deuterium content by LC-MS/MS using the internal standard (trimethyl-d_9)-choline. Due to the poor solubility of choline in organic solvents, we verified only the stereochemical purity of the deuterated ethanolamine precursors. We achieved this by ^{1}H NMR, after derivatization with (1*S*)-(–)-camphanic chloride to afford the N,O-dicamphanoyl product, as previously done for the deuterated etanolamine analogs that were tested with EAL (**Figure 3.25**) (38). Using this assay, we could not detect the minor enantiomer, and thus concluded that we were able to access both (*R*)-[1-^{2}H]-choline and (*S*)-[1-^{2}H]-choline in >95% ee.

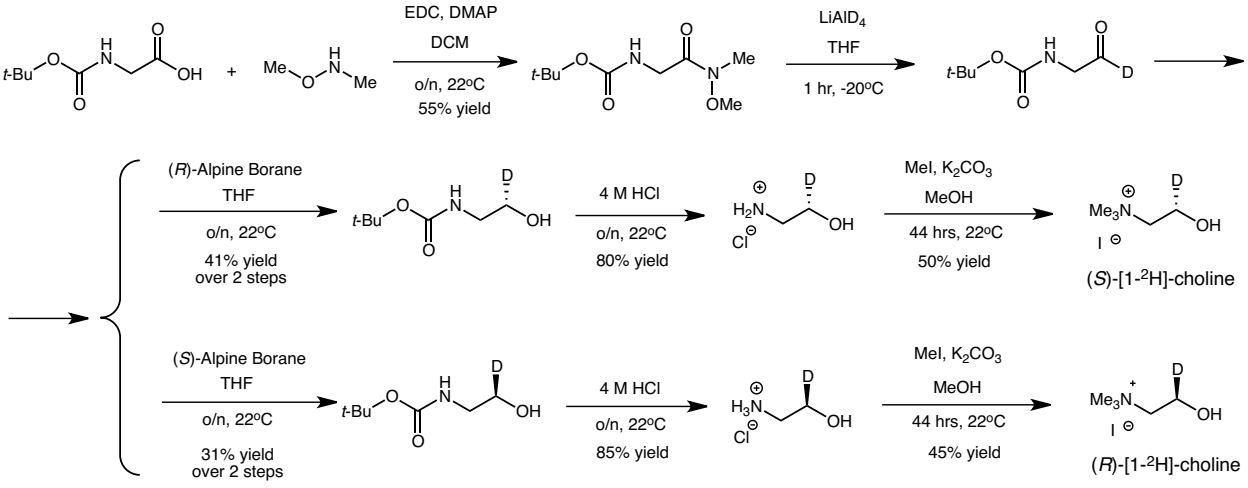


Figure 3.24: Synthetic scheme for accessing (*R*)-[1-^{2}H]-choline and (*S*)-[1-^{2}H]-choline.

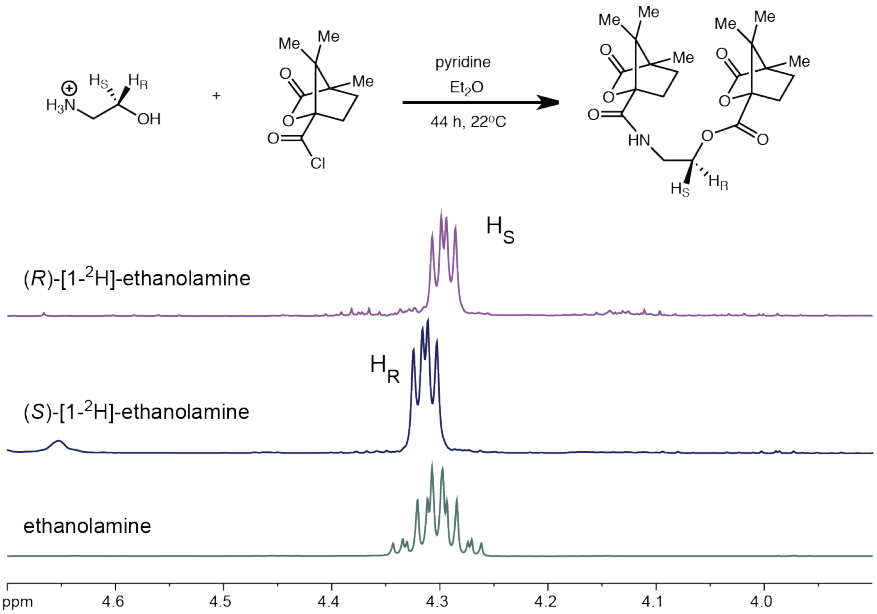


Figure 3.25: NMR assay to verify the stereochemical purity of deuterium labeled ethanolamine analogs.

We tested the stereochemical outcome of the initial hydrogen atom abstraction from choline by performing preliminary deuterium nuclear magnetic resonance (^{2}H NMR) experiments. We incubated stereospecifically deuterated choline analogs with activated CutC for 2 hours (**Figure 3.26**). Since this was an end point assay, we were unable to determine the kinetic parameters of these substrates. The ^{2}H NMR spectra indicated that the deuterium atom from (*R*)-[1-^{2}H]-choline is exclusively retained at C1, while the D atom from (*S*)-[1-^{2}H]-choline is distributed between C1 and C2 over the course of the reaction (**Figure 3.26**). These experiments must be interpreted with caution as they are still preliminary, and the reaction does not appear to have gone to completion for any of the substrates. Nonetheless, one explanation for this apparent lack of stereospecificity of CutC could be that pro-*S* hydrogen atom abstraction from C1 of choline is much faster than pro-*R* hydrogen atom abstraction, but that a large deuterium isotope effect allows the pro-*R* hydrogen atom abstraction to become more dominant in the pro-*S* labeled substrate. Further work (*e.g.,* running the ^{2}H NMR experiments until complete substrate consumption, and determining the kinetic parameters of CutC-catalyzed cleavage of deuterated choline analogs) will be required in order to verify the stereochemical course of this hydrogen abstraction step by CutC.

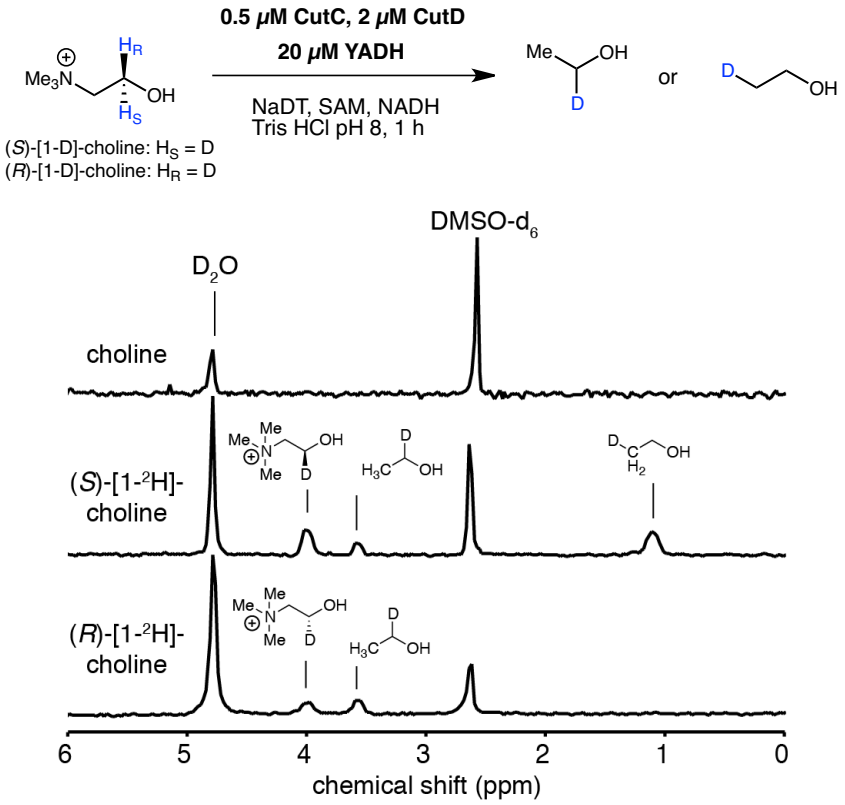


Figure 3.26: Preliminary investigation of the stereochemistry of C1 hydrogen abstraction by CutC. Deuterium-labeled and unlabeled choline were reacted with activated CutC, and the resulting acetaldehyde was reduced by yeast alcohol dehydrogenase (YADH)/NADH. The resulting labeled ethanol was detected by ^{2}H NMR (400 MHz).

3.2.10: Conclusions

In summary, we successfully reconstituted the activity of CutC, the first C–N bond cleaving GRE, and gained initial insights into the basis for its substrate specificity and catalytic activity. Understanding the relationship between the amino acid sequence of CutC and its function enables the use of this enzyme as a biomarker for anaerobic choline metabolism. As we have detailed in Chapter 1, the identification of unique active site residues allowed us to distinguish the sequences of CutC homologs identified in genomic sequencing data from those encoding other GREs, which facilitated the characterization of choline metabolism in microbial communities, including the human gut microbiota. Ultimately, this biochemical knowledge also helps to access CutC inhibitors, as we will present in Chapter 5. In addition to providing valuable mechanistic information, such molecules would also be

important tool compounds for understanding the role of anaerobic choline metabolism in the human gut and its connection to disease.

3.3. Materials and methods

3.3.1: Materials and general methods

All chemicals and solvents were purchased from Sigma-Aldrich, except where otherwise noted. Luria-Bertani Lennox (LB) medium was purchased from EMD Millipore, d_9-trimethylamine (TMA) from CDN Isotopes, choline chloride-1-^{13}C from Sigma-Aldrich, and d_4-acetaldehyde and NMR solvents from Cambridge Isotope Laboratories. Methanol, acetonitrile, and water used for liquid chromatography-mass spectrometry (LC-MS) were B&J Brand high-purity solvents (Honeywell Burdick & Jackson). Oligonucleotide primers were synthesized by Integrated DNA Technologies. Recombinant plasmid DNA was purified with a Qiaprep Kit from Qiagen. Gel extraction of DNA fragments and restriction endonuclease clean up were performed using an Illustra GFX PCR DNA and Gel Band Purification Kit from GE Healthcare. DNA sequencing was performed by Genewiz or Beckman Coulter Genomics. Multiple sequence alignments were performed with Clustal W (39), and the results analyzed with Geneious Pro 6.1.6 (40).

Nickel-nitrilotriacetic acid agarose (Ni-NTA) resin was purchased from Qiagen. SDS-PAGE gels were purchased from BioRad. Protein concentrations were determined according to the method of Bradford using bovine serum albumin (BSA) as a standard (41), or using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific) for CutC (ε = 128 870 M^{-1} cm^{-1} for all CutC variants from *D. alaskensis*, ε = 128 167 M^{-1} cm^{-1} for CutC–full length from *E. coli* MS 69-1, ε = 91 684 M^{-1} cm^{-1} for CutC–334 aa truncation from *E. coli* MS 69-1). The extinction coefficients for CutC were obtained using the ExPASy protparam tool (http://www.expasy.org/tools/protparam.html).

All anaerobic experiments were conducted in an MBraun glovebox (MBraun) under an atmosphere consisting of 99.997% N_2 with less than 5 ppm O_2, or in an anaerobic chamber (Coy Laboratory Products) under an atmosphere of 98% nitrogen and 2% hydrogen. Gel filtration experiments were performed on a BioLogic DuoFlow Chromatography System (Bio-Rad) equipped with a Superdex

200 10/300 GL column (GE Healthcare). Bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa), and vitamin B_{12} (1350 Da) were used as molecular weight markers (Bio-Rad, #151-190). The molecular weights of the proteins analyzed by gel filtration were calculated from their elution volume, using a second-degree polynomial for the relationship between log(molecular weight) and retention time.

Circular dichroism (CD) spectra were taken with a Jasco J-710 Spectropolarimeter in the wavelength range 185–300 nm. The scan speed was 50 nm min^{-1}, the data pitch 0.1 nm, the bandwidth 0.5 nm, and four scans were accumulated per sample. The spectropolarimeter was equipped with a Jasco PTC-348W temperature controller set to 22 °C. The cuvettes used for CD measurements were Type 21 quartz 0.4 mL cells with a PTFE stopper (1 mm path length, Starna Cells; 21-Q-1).

Perpendicular mode X-band electron paramagnetic resonance (EPR) spectra were recorded on a Bruker ElexSysE500 EPR instrument fitted with a quartz dewar (Wilmad Lab-Glass) for measurements at 77 K, or a cryostat for measurements at 9–40 K. All samples were loaded into EPR tubes with 4 mm outer-diameter and 8" length (Wilmad Lab-Glass, 734-LPV-7). Data acquisition was performed with Xepr software (Bruker). The magnetic field was calibrated with a standard sample of α,γ-bisdiphenylene-β-phenylallyl (BDPA), g = 2.0026 (Bruker). The experimental spectra for the glycyl radical were modeled with EasySpin for Matlab to obtain g values, hyperfine coupling constants, and line widths (42). Spin concentration measurements were performed for CutC activation assays by numerically calculating the double integral of the simulated spectra and comparing the area with that of a $K_2(SO_3)_2NO$ standard, without considering the difference in receiver gain because it was already accounted for by the EPR spectrometer. This standard is not stable for a long time (43), so it was prepared before each set of EPR measurements. Solid $K_2(SO_3)_2NO$ was dissolved under anaerobic conditions in anoxic 0.5 M $KHCO_3$ and diluted to a final concentration of 0.3–0.5 mM. To account for any decomposition during dissolution, the concentration was measured at 248 nm (ε = 1690 M^{-1} cm^{-1}) (43), using a NanoDrop 2000 UV-Vis Spectrophotometer. The standard used for quantification of the signal arising from the iron-sulfur clusters of CutD was prepared by combining 800 μM $CuSO_4$ with 8 mM EDTA. For glycyl radical-containing

samples and the $K_2(SO_3)_2NO$ standard, EPR spectra represent the average of 16 scans and were recorded under the following conditions: temperature, 77 K; center field, 3370 Gauss; range, 200 Gauss; microwave power, 20 μW; microwave frequency, 9.45 MHz; modulation amplitude, 0.4 mT; modulation frequency, 100 kHz; time constant, 20.48 ms; conversion time, 20.48 ms; scan time, 20.97 s; receiver gain, 60 dB (for enzymatic assays) or 30 dB (for standards). For iron-sulfur cluster-containing samples and the Cu-EDTA standard, EPR spectra represent the average of 4 scans, and were recorded under similar conditions except for: temperature, 9 – 40 K, center field, 3500 Gauss; range, 2000 Gauss; microwave power, 2 mW.

Analytical HPLC was performed on a Dionex Ultimate 3000 instrument (Thermo Scientific). The spectrophotometric coupled assay for CutC activity was conducted at 20 ºC in 96-well plates using a PowerWave HT Microplate Spectrophotometer (BioTek). The absorbance of each well was monitored at 340 nm for 10 min, with reads 20 s apart. Path length correction was employed such that absorbance values were reported for a path length of 1 cm. Data was analyzed using Gen5 Data Analysis Software and kinetic parameters were calculated using GraphPad Prism6. The UV-Vis spectra for CutD were collected using the same instrument and a quartz 96-well plate.

LC-MS/MS analysis was performed in our research lab in the Department of Chemistry and Chemical Biology, Harvard University, on an Agilent 6410 Triple Quadrupole LC/MS instrument (Agilent Technologies), or in the Small Molecule Mass Spectrometry Facility at Harvard University on an Agilent 6460 Triple Quadrupole Mass Spectrometer with Agilent 1290 uHPLC (Agilent Technologies). The same LC-MS/MS method was used as previously described for the quantification of TMA (2), with the exception that for the Agilent 6460 Triple Quadrupole Mass Spectrometer, the drying gas temperature was maintained at 350 °C with a flow rate of 12 L min^{-1} and a nebulizer pressure of 25 psi. For the analysis of stereospecifically deuterated choline, the precursor-product ion pairs used in MRM mode were: m/z 113.2→m/z 69.2 (trimethyl-d_9-choline), m/z 106.1→m/z 60.2 ([1-$^2H_2$]-choline), m/z 105.1→ m/z 60.2 ([1-^{2}H]-choline), m/z 104.1→m/z 60.2 (unlabeled choline), and the collision energy was 17 V. Liquid chromatography was performed as described in Chapter 2.

All GC-MS experiments were conducted using a Waters Quattro micro GC Mass Spectrometer (Waters) equipped with a Combi PAL headspace autosampler (CTC Analytics) and a split/splitless injector. The column used was a DB-624UI, 30 m x 0.32 mm x 1.80 µm (Agilent) and the inlet liner was a 1 mm straight single taper Ultra Inert liner (Agilent). The carrier gas was helium, held at 2.3 mL min^{-1} constant flow. Samples were incubated at 37 °C for 30 min, with an agitator speed of 500 rpm. The GC injection port was set at 220 °C, the syringe temperature to 120 °C, and the needle flush time was 120 s. A headspace sample volume of 1 mL was injected into the inlet with a split ratio of 20:1. The column temperature was initially maintained at 30 °C for 3 min, then increased to 250 °C at a rate of 50 °C min^{-1} with a total run time of 7.4 min. The GC was coupled to a mass spectrometer, equipped with an electron ionization source and a triple quadrupole mass analyzer with acquisition in positive mode. The GC interface temperature was maintained at 220 °C, and the ion source temperature at 200 °C. In order to identify the products present in enzymatic reactions, the mass analyzer was set to scan over a 45–100 m/z range for 0–1.3 min and a 20–100 m/z range for 1.3–6.2 min of the run. The Selected Ion Recording (SIR) function was used for detection and quantification of acetaldehyde over the first 6.2 min of the run. The monitored SIR ions were m/z 29 for unlabeled acetaldehyde and m/z 30 for d_4-acetaldehyde. GC-MS data analysis was performed with the MassLynx software package. Peak intensities for the ions indicated above, at a retention time of ~1.44 min, were used for the final quantification of acetaldehyde.

Proton (^1H), deuterium (^2H), and carbon (^{13}C) nuclear magnetic resonance (NMR) spectra were recorded in the Magnetic Resonance Laboratory in the Harvard University Department of Chemistry and Chemical Biology on a Varian Inova-500 (500 MHz, 125 MHz) or Varian Mercury-400 (400 MHz, 100 MHz) NMR spectrometer. Chemical shifts are reported in parts per million downfield from tetramethylsilane using the solvent resonance as an internal standard for ^1H (D_2O = 4.75 ppm), and ^{13}C (DMSO-d_6= 39.52 ppm), and a small amount of deuterated solvent for ^2H NMR (DMSO-d_6 in H_2O = 2.71 ppm). Data are reported as follows: chemical shift, integration multiplicity (s = singlet, d = doublet, t = triplet), coupling constant, integration, and assignment. NMR spectra were visualized using ACD/NMR Processor Academic Edition.

High-resolution mass spectral (HRMS) data for the synthetic compounds were obtained in the Magnetic Resonance Laboratory in the Harvard University Department of Chemistry and Chemical Biology on a Bruker Micro QTOF-QII fitted with a dual-spray electrospray ionization (ESI) source. The capillary voltage was set to 4.5 kV and the end plate offset to –500 V, the drying gas temperature was maintained at 190 ^{o}C with a flow rate of 8 L min^{-1} and a nebulizer pressure of 21.8 psi. The liquid chromatography (LC) was performed using an Agilent Technologies 1100 series LC with 50% H_2O and 50% acetonitrile as solvent.

3.3.2: Cloning of cutC (wild type, truncated variants, and mutants) and cutD

Table 3.3: Oligonucleotides used for cloning. Restriction sites are underlined.

Oligonucleotide	Target	Sequence (5' to 3')	Restriction site
For cloning *cutC* and *cutD* from *Desulfovibrio alaskensis* G20			
Dde-3282-NdeI-start	CutC	GCATCATATGGATCTCCAGGACTTTTCACATAAGC	NdeI
Dde-3282-XhoI-stop	CutC CutC (–18 aa) CutC (–30 aa)	GATCCTCGAGTTAGAAACCATGCAGCATGG	XhoI
CutC-truncated-18aa-NdeI-start	CutC (–18 aa)	GCATCATATGCCGGCAGAACGTGCTTCG	NdeI
CutC-truncated-30aa-NdeI-start	CutC (–30 aa)	GCATCATATGGGCGTATCCGCCGAGGTCTTC	NdeI
CutC-truncated-52aa-NdeI-start	CutC (–52 aa)	GCATCATATGGGCATTCCCGACGGGCCTAC	NdeI
Ddes3282-NotI-stop	CutC (–52 aa)	GATTAA GCG GCC GCTTAGAAACCATGCAGC	NotI
Dde-3281-NdeI-start	CutD	GCATCATATGAGAACCGCAACACACAGAGACG	NdeI
Dde-3281-XhoI-nostop	CutD	GATTCTCGAGGTGGCGGATCACCGAAACC	XhoI
Dde3282-QC-D216N-for	CutC-D216N	GTGAACGGCGGCGGCAACTCCAACCCCGGTTAC	–
Dde3282-QC-D216N-rev	CutC-D216N	GTAACCGGGGTTGGAGTTGCCGCCGCCGTTCAC	–
Dde3282-QC-T334S-for	CutC-T334S	GTTGAAGAAAACCAGAGCGGTATGTCCATCGGC	–
Dde3282-QC-T334S-rev	CutC-T334S	GCCGATGGACATACCGCTCTGGTTTTCTTCAAC	–
Dde3282-QC-F395L-for	CutC-F395L	GCCGGGTACCAGCCTTTAGTGAACATGTGCGTG	–
Dde3282-QC-F395L-rev	CutC-F395L	CACGCACATGTTCACTAAAGGCTGGTACCCGGC	–
Dde3282-QC-C489A-for	CutC-C489A CutC (–52 aa) C489A	GACTACTGCCTGATGGGTGCCGTGGAACCGCAG	–
Dde3282-QC-C489A-rev	CutC-C489A	CTGCGGTTCCACGGCACCCATCAGGCAGTAGTC	–
Dde3282-QC-E491Q-for	CutC-E491Q	CTGATGGGTTGCGTGCAACCGCAGAAATCAGGCCGT	–
Dde3282-QC-E491Q-rev	CutC-E491Q	ACGGCCTGATTTCTGCGGTTGCACGCAACCCATCAG	–
Dde3282-QC-T502S-for	CutC-T502S	CGTCTGTACCAGTGGAGCTCCACCGGCTATACC	–
Dde3282-QC-T502S-rev	CutC-T502S	GGTATAGCCGGTGGAGCTCCACTGGTACAGACG	–

Dde3282-QC-G821A-for	CutC-G821A	GTGGTGCGCGTGGCCGCATACAGCGCCTTCTTC	–
Dde3282-QC-G821A-rev	CutC-G821A	GAAGAAGGCGCTGTATGCGGCCACGCGCACCAC	–
For cloning *cutC* from *Escherichia coli* MS 69-1			
MS69CutCtrunc334aa-NdeI-start	CutC –334 aa	GAGCGTCATATGGTTACCACACTATGTGCCGCC	NdeI
MS69CutC-Chis-NotI-nostop	CutC CutC –334 aa	GAATTAGCGGCCGCGAACTTCTCAATCACCGTAC	NotI
MS69CutC-NdeI-start	CutC	GCCCCGCATATGCCACTGATAACAAGAACTGG	NdeI

Amplification of *cutC* (Dde_3282) and *cutD* (Dde_3281) from *Desulfovibrio alaskensis* G20 genomic DNA, and of *cutC* (EFJ81575.1) from *Escherichia coli* MS 69-1 genomic DNA were performed according to a previously published procedure (2) using the primers listed in **Table 3.3**. Amplified fragments were digested with the appropriate restriction enzymes (*NdeI, NotI, XhoI*) *(*New England Biolabs, NEB), purified and ligated into linearized expression vectors as previously described (2), to afford pET-29b-CutD (C-terminal His_6-tagged construct), pET-28a-CutC from *D. alaskensis* G20 (N-terminal His_6-tagged constructs), and pET-29b-CutC-full length and pET-29b-CutC-334aa from *E. coli* MS 69-1 (C-terminal His_6-tagged constructs). These constructs were transformed into chemically competent *E. coli* BL21 (DE3) cells (Invitrogen) and stored at −80 °C as frozen LB/glycerol stocks.

Site-directed mutagenesis of CutC (full length enzyme from *D. alaskensis* G20) was performed using the corresponding oligonucleotides listed in **Table 3.3**. PCR reactions of 20 μL contained 10 μL of Phusion High-Fidelity PCR Master Mix (New England Biolabs), 50 ng of pET-28a-CutC template, and 100 pmoles of each primer. Thermocycling was carried out in a C1000 Gradient Cycler (Bio-Rad) using the following parameters: denaturation for 1 min at 95 °C, followed by 18 cycles of 30 s at 95 °C, 1 min at 67 °C, and a final extension time of 8 min at 72 °C. Digestion of the methylated template plasmid was performed with DpnI (NEB), and 5 μL of each digestion was used to transform a single tube of chemically competent *E. coli* TOP10 cells. The identities of the resulting pET-28a-CutC-D216N, pET-28a-CutC-T334S, pET-28a-CutC-F395L, pET-28a-CutC-C489A, pET-28a-CutC-E491Q, pET-28a-CutC-T502S, pET-28a-CutC-G821A, and pET-28a-CutC-52-C489A constructs were confirmed by sequencing

purified plasmid DNA. Chemically competent *E. coli* BL21(DE3) cells were transformed with plasmid DNA and stored at –80 °C as frozen LB/glycerol stocks.

3.3.3: Large scale overexpression and purification of wild type CutC, truncated variants, and active site mutants

An LB agar plate with 50 μg mL^{-1} kanamycin (Kan) was streaked with a frozen stock of *E. coli* BL21(DE3) cells transformed with a CutC-harboring plasmid. A single colony from each plate was inoculated into 50 mL of LB medium with 50 μg/ml Kan, which was grown overnight at 37 °C. These starter cultures were diluted 1:100 into 2 L of LB medium containing 50 μg mL^{-1} Kan, then incubated at 37 °C with shaking at 175 rpm, induced with 500 μM IPTG (Teknova) at OD_{600} = 0.5–0.6, and incubated at 37 °C for 4–5 h (wild type CutC and active site mutants from *D. alaskensis* G20) or at 15 °C overnight (truncated CutC variants from *D. alaskensis* G20, CutC–full length and –334 aa truncation from *E. coli* MS 69-1). Cells from 2 L of culture were harvested by centrifugation (6,000 rpm x 10 min), resuspended in 45-50 mL of lysis buffer. The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 8,000–10,000 psi, and the lysate was clarified by centrifugation (10,000 rpm x 30 min). The supernatant was incubated with 2 mL of Ni-NTA resin and 5 mM imidazole for 2 h (30 min for CutC from *E. coli*) at 4 °C. The mixture was centrifuged (3,000 rpm x 5 min) and the unbound fraction discarded. The Ni-NTA resin was resuspended in 10 mL of elution buffer, loaded into a glass column, and washed with 10 mL of elution buffer. Protein was eluted from the column using a stepwise imidazole gradient in elution buffer (25 mM, 50 mM, 100 mM, 200 mM, 500 mM), collecting 2 mL fractions. SDS–PAGE analysis (4–15% Tris-HCl gel) was employed to ascertain the presence and purity of protein in each fraction. Fractions containing the desired protein were combined and dialyzed twice against 2 L of storage buffer. Upon removal of any precipitated protein through centrifugation at 3,000 rpm x 2 min, the solutions were concentrated using a Spin-X® UF 20 mL centrifugal concentrator with a 30,000 MWCO membrane (Corning®), then frozen in liquid N_2 and stored at –80 °C.

All CutC mutants and the preparations of CutC WT used for MS assays from *D. alaskensis* G20 were purified using this protocol and the following buffers: lysis buffer contained 20 mM Tris-HCl pH 8,

500 mM NaCl, 10 mM MgCl2, and half of a SIGMAFAST™ protease inhibitor cocktail tablet (Sigma-Aldrich), elution buffer consisted of 20 mM Tris-HCl pH 8, 500 mM NaCl, 10 mM $MgCl_2$, and 5 mM imidazole, and storage buffer contained 20 mM Tris-HCl pH 8, 50 mM NaCl, 10% (v/v) glycerol. This procedure afforded yields of 1–3 mg L^{-1} of bacterial culture for CutC WT, 2 mg L^{-1} for CutC-D216N, 6 mg L^{-1} for CutC-T334S, 16 mg L^{-1} for CutC-F395L, 1.8 mg L^{-1} for CutC-C489A, 0.9 mg L^{-1} for CutC-E491Q, 9 mg L^{-1} for CutC-T502S, and 1.5 mg L^{-1} for CutC-G821A (**Figure 3.27**). Proper folding of CutC WT and mutant proteins was verified by analyzing a 10 μM solution of each protein by circular dichroism, as outlined in 'materials and general methods'.

Since the yields obtained by purifying CutC WT with Tris-HCl pH 8 buffer were too low to allow its characterization by EPR, we optimized a purification procedure employing potassium phosphate buffer pH 8. This protocol was also used to purify the truncated variants of CutC from *D. alaskensis* G20: CutC –18 aa, CutC–30 aa, CutC–52 aa, the mutant CutC–52aa C489A, and CutC–full length and –334 aa truncated variant from *E. coli* MS 69-1. The overall procedure was similar to the one described above with the following exceptions: the lysis buffer contained 50 mM potassium phosphate pH 8 and 300 mM KCl; no protease inhibitor was added; the elution buffer contained 50 mM potassium phosphate pH 8, 300 mM KCl, and 5 mM imidazole; and the storage buffer consisted of 50 mM potassium phosphate pH 8, 50 mM KCl, and 10% (v/v) glycerol. For CutC from *D. alaskensis* G20, this procedure afforded yields of 6-10 mg L^{-1} of bacterial culture for CutC WT, 17 mg L^{-1} for CutC (–18 aa), 20 mg L^{-1} for CutC (–30 aa), 24 mg L^{-1} for CutC (–52 aa), and 25 mg L^{-1} for CutC (–52 aa) C489A. For CutC from *E. coli* MS 69-1, this procedure afforded yields of 20 mg L^{-1} of bacterial culture for CutC–full length and 17 mg L^{-1} for CutC–334 aa truncation.

3.3.4: Determination of the molecular mass of CutC using size exclusion chromatography

A ~30 μM solution of CutC (full-length or truncated variant) was analyzed by size exclusion chromatography, as described in 'materials and general methods' (50 μL injection volume). The proteins were eluted over 122 min with 50 mM potassium phosphate buffer pH 8, 50 mM KCl, 10% v/v glycerol

at 0.25 mL min^{-1}. A solution of molecular weight markers (Bio-Rad, #151-190) analyzed under the same conditions and was used to construct a standard curve.

3.3.5: Large scale overexpression and purification of CutD

The best CutD yields were achieved when plasmid pPH149 (21) containing the *E. coli IscSUA-HscBA-Fd* genes (ISC system) was co-transformed into chemically competent *E. coli* BL21 (DE3) cells (Invitrogen) together with pET-29b-CutD. An LB agar plate with 50 μg mL^{-1} Kan and 50 μg mL^{-1} chloramphenicol (Cam) was streaked with a frozen stock of *E. coli* BL21(DE3) cells transformed with pET-29b-CutD and the ISC system. A single colony was used to inoculate 50 mL of LB medium with 50 μg mL^{-1} Kan and 50 μg mL^{-1} Cam, which was grown overnight at 37 °C. This starter culture were diluted 1:100 into a screw-capped 2.8 L flask with 2 L of LB medium supplemented with 50 μg mL^{-1} Kan, 50 μg mL^{-1} Cam, 2 mM iron (III) ammonium citrate, and 0.5 % (w/v) glucose. The culture was incubated at 37 °C with shaking at 175 rpm until reaching an OD_{600} = 0.35–0.4, at which point it was placed at room temperature and sparged with argon gas for ~30 min. Upon addition of L-cysteine (2 mM final concentration) and sodium fumarate (20 mM final concentration), the culture was further sparged until reaching an OD_{600} = 0.5–0.6, at which point protein expression was induced with 500 μM IPTG (Teknova), the flask was capped tightly and wrapped with electrical tape, and the culture was incubated at 15 °C overnight (~16 to 18 h) without shaking in an Mbraun glove box.

All purification steps were performed in an anaerobic chamber at 11 °C (MBraun glove box), with the exception of cell lysis and centrifugation. Cells from 2 L of culture were harvested by centrifugation in 250 mL polypropylene centrifuge tubes (6,000 rpm x 10 min) and resuspended in 45-50 mL of anoxic lysis buffer (50 mM HEPES, 200 mM NaCl, 10 mM $MgCl_2$, pH 8) supplemented with half of a SIGMAFAST™ protease inhibitor cocktail tablet (Sigma-Aldrich), 8 mg of chicken egg lysozyme, and 5 mM DL-dithiothreitol (DTT). The resuspension was brought out of the anaerobic chamber and the cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 8,000–10,000 psi while maintaining a nitrogen line above the receiving tube and one above the cell disruptor reservoir. For assays with CutC from *E. coli* MS 69-1, CutD-expressing cells were lysed under strict anaerobic

conditions, in a Coy Labs glove box at 4 °C, employing a sonifier (Branson). The lysate was clarified by centrifugation (10,000 rpm x 30 min), and the supernatant was incubated with 2 mL of Ni-NTA resin (previously sparged with nitrogen) and 5 mM imidazole for 2 h at 4 °C. The mixture was brought out of the anaerobic chamber, centrifuged (3,000 rpm x 5 min), and the unbound fraction discarded. The Ni-NTA resin was resuspended inside the anaerobic chamber in 10 mL of elution buffer (50 mM HEPES, 200 mM NaCl, 10 mM $MgCl_2$, 5 mM imidazole, pH 8) and loaded into a glass column. Protein was eluted from the column using a stepwise imidazole gradient in elution buffer (10 mM, 25 mM, 50 mM, 100 mM, 200 mM), collecting 2 mL fractions. SDS–PAGE analysis (4–15% Tris-HCl gel) was employed to ascertain the presence and purity of protein in each fraction. Fractions containing CutD were combined and stirred slowly for 13–16 h with 2 mM DTT, 0.2 mM Na_2S, and 0.25 mM $Fe(NH_4)_2(SO_4)_2$ (the iron was added from a 10 mM stock 5 min after the DTT and Na_2S solutions were added).

Reconstituted CutD was filtered through a 25 mm, 0.22 μm pore-size Acrodisc syringe filter with HT Tuffryn Membrane (Pall Life Sciences) to remove particulates and was dialyzed with three 1 L batches of 50 mM HEPES pH 8, 50 mM NaCl, and 10 % (v/v) glycerol. Upon concentration, the protein solution was aliquoted into 0.5 mL cryogenic vials, frozen in liquid N_2, and placed in 18 x 150 mm Hungate tubes (Chemglass) sealed with butyl stoppers and aluminum seals. This procedure afforded up to 9 mg of CutD per L of bacterial culture (**Figure 3.27**), while expression without the ISC system provided a 2-fold lower yield, without impacting activity.

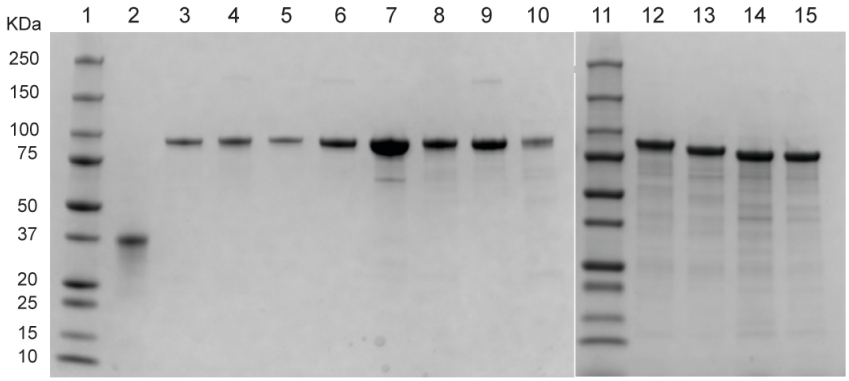


Figure 3.27: SDS-PAGE (4-15% Tris-HCl gel) of CutC and CutD from *D. alaskensis* G20. Precision Plus Protein™ All Blue Standards (Bio Rad) (lanes 1 and 11), CutD (lane 2), CutC WT (lane 3), CutC mutants: CutC-D216N (lane 4), CutC-T334S (lane 5), CutC-F395L (lane 6), CutC-C489A (lane 7), CutC-

E491Q (lane 8), CutC-T502S (lane 9), CutC-G821A (lane 10), CutC truncated variants: CutC–18 aa (lane 12), CutC–30 aa (lane 13), CutC–52 aa (lane 14), CutC–52 aa-C489A (lane 15).

3.3.6: Overexpression and purification of flavodoxin and flavodoxin reductase from E. coli *K-12*

The plasmids for flavodoxin (pET-23b-Fld) and flavodoxin reductase (pET-28b-Fpr) from *E. coli* K-12 were constructed as previously described (44, 45). Overexpression and purification of these two enzymes employed a published method (45) with the following modifications: 10 μM of riboflavin was added to the growth medium prior to induction with IPTG, and the storage buffer used was 20 mM sodium phosphate pH 7.4, 50 mM NaCl, 10 % (v/v) glycerol. We obtained yields of 28 mg L^{-1} of bacterial culture for Fld, and 60 mg L^{-1} for Fpr.

3.3.7: UV-Vis assay for assembly of the [4Fe-4S] cluster in CutD

After reconstitution of CutD, the protein solution was diluted inside the anaerobic chamber to 10 μM with anoxic buffer (50 mM HEPES pH 8, 50 mM NaCl), and its absorbance was measured over a 250–800 nm range using a quartz 96-well plate as described in 'materials and general methods'. To obtain the spectrum for reduced CutD, 10 μM of the protein was incubated with a 10-fold excess of sodium dithionite (NaDT) for 15 min prior to the absorbance measurement.

3.3.8: Determination of iron content of CutD

The iron content of a 6 μM solution of CutD was determined using Ferene (3-(2-pyridyl)-5,6-di(2-furyl)-1,2,4-triazine-5 and 5''-disulfonic acid disodium salt), according to a previously published procedure (46), with the exceptions that the assay volume was tripled and the standard curve was prepared with ammonium iron (II) sulfate hexahydrate (0-70 μM).

3.3.9: Determination of sulfide content of CutD

The sulfide content of a 6 μM solution of CutD was measured using a previously published procedure (47), with the following modifications: the assays were performed in microcentrifuge tubes, vortexing was used instead of stirring, and the mixture was incubated for 20 min at room temperature after addition of NaOH.

3.3.10: EPR characterization of CutD

The samples for EPR analysis of the reduced iron-sulfur cluster contained 50 mM HEPES pH 8, 50 mM NaCl, 40 μM CutD, and 4 mM NaDT, and were incubated at room temperature (22 °C) under anaerobic conditions for 20 min prior to freezing in liquid N_2. In order to determine the effect of S-adenosylmethionine (SAM) on the EPR signal, 4 mM SAM was added after the reduction, and the sample was frozen within less than 5 min. Quantification of the signal at 9 K for each sample was performed with a Cu-EDTA standard as described in the 'materials and general methods', and indicated the presence of ~0.8 spins per CutD monomer in each case. Despite the apparent splitting noticeable in the (A) and (B) spectra from **Figure 3.8** (which could be due to a superimposition of signals), the signal with g|| of 2.045 and $g\perp$ of 1.94 can be unambiguously attributed to a $[4Fe\text{-}4S]^{+}$ cluster, because this signal broadens and decreases in intensity at higher (40 K) temperatures, as shown in spectrum (C). The shape and amplitude of the EPR signal do not seem to change upon addition of excess SAM, an observation that has been made for other glycyl radical activating proteins (27, 48, 49).

3.3.11: Detection of 5'-deoxyadenosine by HPLC

The 5'-dA detection assay contained 25 mM Tris HCl buffer pH 8, 50 mM NaCl, 16 mM NaDT, 40 μM CutD, and 1 mM SAM in a total volume of 100 μL. All components except for SAM were combined inside an anaerobic chamber in order to allow CutD to reduce and SAM was added after 15 min. The reactions were incubated at room temperature in the chamber for 4 h, quenched with 2.56 μL formic acid, and centrifuged (13,200 rpm x 10 min) to remove precipitated protein. The supernatant was analyzed by HPLC (20 μL injection volume) on a Phenomenex Luna analytical C18 reverse phase column (5 μm, 4.6 x 250 mm, 100 Å), preceded by a HAIPEEK guard cartridge holder (Higgins Analytical) equipped with a CLIPEUS C18 (5 μm, 2 cm x 3.2 mm) cartridge (Higgins Analytical). The solvent system consisted of 40 mM ammonium acetate pH 6 (A) and acetonitrile (B), and the elution conditions were: 0% B for 2 min, 0-25% B over 25 min, 25-0% B over 2 min, 0% B for 5 min. The flow

rate was set to 1 mL min^{-1}, the products were detected by UV at 216 nm and 260 nm, and commercial 5'-deoxyadenosine was used as a standard.

3.3.12: Detection of methionine by HPLC

The methionine detection assay was performed in a similar manner as the 5'-dA detection assay, except that it contained 12 mM NaDT and 30 μM CutD, and it was incubated for 18 h. The reactions were quenched with 200 μL methanol, kept on ice for 10 min, and centrifuged (13,200 rpm x 10 min). The organic solvent was removed *in vacuo*, concentrating the mixtures to ~100 μL. A 20 μL aliquot of each assay was mixed with o-phthaldialdehyde (OPA) reagent (Sigma-Aldrich; containing 2 μL of 2-mercaptoethanol per mL of incomplete reagent). After an incubation of 5-6 min, the samples were analyzed by HPLC (40 μL injection volume) on a Dionex Acclaim C18 reverse phase column (3 μm, 2.1 x 150 mm, 120 Å), preceded by the same guard cartridge as for the 5'-dA assay. The solvent system consisted of water (A) and acetonitrile (B). The products eluted at 0.25 mL min^{-1} with the following elution conditions: 0% B for 1 min, 0-70% B over 15 min, 70-0% B over 5 min, 0% B for 2 min. OPA-derivatized L-methionine was used as a standard and the products were detected by UV at 338 nm.

3.3.13: Quantification of the glycyl radical content of CutC truncated variants from D. alaskensis *G20*

All activation assays contained 25 mM Tris HCl buffer pH 8, 50 mM NaCl, 3 mM NaDT, 30 μM CutD, 7.5 μM CutC dimeric truncated variant (–18 aa, –30 aa, or –52 aa), and 3 mM SAM in a total volume of 300 μL. For all assays involving CutC, the enzyme was made anoxic by sparging with argon for 15-30 min. All components except for CutC and SAM were combined inside an Mbraun anaerobic chamber in order to allow CutD to reduce, then SAM was added after 20 min. The reactions were incubated at room temperature for approximately 60 min before being transferred to an EPR tube and frozen in liquid nitrogen. For each CutC truncated variant, the activation assay was performed in triplicate. Quantification of the amount of glycyl radical present in the assay was done with a $K_2(SO_3)_2NO$ standard as described in the 'general materials and methods'. The extent of activation was calculated for each variant and is reported as % of CutC polypeptides converted to the glycyl radical form: 8.2 ± 0.2 %

for CutC (–18 aa), 8.6 ± 0.3 % for CutC (–30 aa), 9.3 ± 0.4 % CutC (–52 aa). Interestingly, for PFL (50) and class III RNR (25), the maximum extent of activation achieved in vitro is 50% of polypeptides, and the explanation for these results still remains unknown. So far, our attempts to increase the extent of CutC activation have been unsuccessful. The amount of glycyl radical observed by EPR was lower when the CutC:CutD ratio employed was 1:0.4 (with either 1 h or 5.5 h incubation), when choline (0.3 mM) was included in the activation assay, and when NaDT was replaced with 1.7 mM DTT, 4 μM Fld, 2 μM Fpr and 0.8 mM NADPH (the NADPH:flavodoxin reductase – flavodoxin system from *E. coli*). Preliminary experiments also indicated that the maximum amount of glycyl radical is obtained after 1 h of incubation and that the radical has a half-life of ~6 h.

3.3.14: Quantification of the glycyl radical content of CutC from E. coli *MS 69-1*

When we analyzed the glycyl radical content of CutC from *E. coli* MS 69-1, we had already carried out extensive efforts to optimize the activation of CutC from *D. alaskensis* G20 (detailed in Chapter 4 of this thesis). As a result, activation assays for CutC from *E. coli* MS 69-1 were set up in an MBraun anaerobic chamber in 1.5 mL polypropylene Eppendorf tubes and contained 50 mM potassium phosphate pH 8, 50 mM potassium chloride, 400 μM sodium dithionite, 400 μM AdoMet, 20 μM CutC dimer and 80 μM CutD from *D. alaskensis* G20 in a total volume of 200 μL. A mixture of buffer, CutD and sodium dithionite was incubated for 20 min prior to addition of AdoMet and CutC, then glycyl radical formation was carried out for 1 h at room temperature. X-band EPR spectroscopy and spin concentration measurements were performed as described in 'materials and general methods'.

3.3.15: Activation assays for CutC from D. alaskensis *G20 conducted in buffered H_2O and D_2O*

In order to exchange CutC–52 aa (wild type or C489A mutant) and CutD from buffered H_2O to D_2O, a 25 mM Tris DCl pH 8, 50 mM NaCl buffer was prepared anoxically in 99.9% D_2O and used to swell 0.5 g of Sephadex G-25 fine resin. A mixture of CutC–52 aa and CutD was passed through the resin, and the proteins were eluted with the D_2O-containing buffer described above. In order to determine which of the 0.2 mL-fractions collected contained protein, 10 μL of each fraction was mixed with 90 μL Bradford reagent. The ones that afforded a color change were pooled and concentrated to ~280 μL with

Spin-X® UF 6 mL centrifugal concentrators with a 5,000 MWCO membrane (Corning®) placed in 50 mL conical-bottom centrifuge tubes with plug seal caps. The activation assays were conducted as presented above and contained 25 mM Tris HCl (or DCl) buffer pH 8, 50 mM NaCl, 5 mM NaDT, 50 μM CutD, 12.5 μM CutC–52 aa (wild type or C489A mutant), and 1 mM SAM in a total volume of 300 μL. EPR spectra were recorded and modeled as described in the 'materials and general methods'.

3.3.16: LC-MS/MS assay for trimethylamine detection from in vitro assays

A typical assay for LC-MS/MS detection of trimethylamine (TMA) contained 25 mM Tris HCl buffer pH 8, 50 mM NaCl, 2 mM NaDT, 5 μM CutD, 0.5 μM CutC (WT, truncated variant, or mutant), substrate (0.2 mM of choline, betaine aldehyde, glycine betaine, DL-carnitine, (*R*)-2-methylcholine, or (*S*)-2-methylcholine), and 1 mM SAM in a total volume of 250 μL. No difference in activity was observed between CutC purified with Tris-HCl pH 8 buffer and with potassium phosphate pH 8 buffer. To a 2 mL screw top amber vial (Agilent Technologies, 5188-6535) located inside the anaerobic chamber were added buffer, NaDT, and CutD, and the resulting mixture was incubated at room temperature for 15 min in order to allow CutD to reduce. CutC, substrate, and SAM were added in this order, the vials were capped immediately with a screw top cap with PTFE/red silicone septa (Agilent Technologies, 5182-071 7), and the reactions were incubated at room temperature for 3 h (exception: the assay with DL-carnitine as substrate was incubated for 13 h). Upon removal from the anaerobic chamber, the TMA concentration in each assay was determined using LC-MS/MS after derivatization with ethyl bromoacetate (2) and using 100 μL of d_9-TMA (200 μM in water) as the internal standard. A 10 μL aliquot of the quenched derivatization mixture was diluted 50-fold with infusion solution [acetonitrile/ water/ formic acid, 50/50/0.025 (v/v/v)] and analyzed by LC-MS (3 μL injection volume) as described in 'materials and general methods'. For the assays that did not afford TMA, the concentration of internal standard used was 0.3 μM, and the samples were not diluted before analysis.

3.3.17: GC-MS/MS assay for acetaldehyde, acetaldehyde-1-^{13}C, propionaldehyde, and 3-hydroxy-propionaldehyde detection

The enzymatic reactions were conducted as described for the LC-MS/MS assays, except that different amounts of substrate were used (1 mM of choline chloride, choline chloride-1-^{13}C, dimethylaminoethanol, methylaminoethanol, ethanolamine, or 1,2-propandiol, and 5 mM of *N*-(2-hydroxyethyl)-*N,N*-dimethylpropan-1-aminium, 3-hydroxy-*N*-(2-hydroxyethyl)-*N,N*-dimethylpropan-1-aminium, *N*-(2-hydroxyethyl)-*N,N*-dimethylbutan-1-aminium), and the incubation time was ~12 h. CutC WT used for these assays was purified with potassium phosphate buffer pH 8. Since both CutC and CutD are stored with 10% (v/v) glycerol (179 mM final concentration), no additional glycerol was added to the assays testing it as a potential substrate. Upon removal from the anaerobic chamber, reaction mixtures were kept at 4 °C for 1 h, uncapped, and the contents were quickly transferred to 10 mL headspace vials with screw tops (Sigma-Aldrich, SU860099) containing 1 mM d_4-acetaldehyde (only for acetaldehyde quantification assays), 1.8 g NaCl, and water up to a final volume of 2.5 mL. The headspace of each vial was analyzed by GC-MS/MS as outlined in 'materials and general methods'. For assays with choline chloride-1-^{13}C, the % ^{13}C incorporation was calculated based on the relative intensities of the fragments with m/z 29 and 30 from the MS/MS spectrum of ^{13}C-acetaldehyde, also taking into consideration the abundance ratio of the fragments with m/z 28 and 29 in the MS/MS spectrum of unlabeled acetaldehyde.

3.3.18: Spectrophotometric coupled assay for kinetic analysis of choline cleavage

Before each assay, CutC–52 aa from *D. alaskensis* G20 was activated without substrate using the same conditions described for the EPR quantification of the glycyl radical. After 1 h of incubation, this activation mixture was diluted 75-fold and added to the kinetic assay. Each coupled assay contained 25 mM Tris HCl buffer pH 8, 50 mM NaCl, 200 μM NADH, 320 nM yeast alcohol dehydrogenase, 8 nM CutC dimer–52 aa, 32 nM CutD, 3.2 μM NaDT, 3.2 μM SAM, and 0-2 mM choline, in a total volume of 200 μL. Activation of CutC from *E. coli* MS 69-1 was carried out as described above. All assays with CutC–full length and CutC–334 aa from *E. coli* MS 69-1 contained 200 μM NADH, 50 mM potassium phosphate pH 8, 50 mM potassium chloride, diluted activation mixture (to achieve a final concentration

of 10 nM CutC dimer), YADH (0.4 μM) and choline (0-2 mM) in a total volume of 200 μL. All reactions were initiated by adding choline. The assays were carried out in triplicate in a 96-well plate and the NADH absorbance at 340 nm was monitored for 5 min. The absorbance decreased linearly in all cases, indicating a constant concentration of activated enzyme during the course of our measurements. The turnover numbers were calculated as the μmoles of product formed per second per μmol of activated CutC (converted to the glycyl radical form).

3.3.19: ^{2}H NMR assays for detection of CutC-mediated production of deuterated acetaldehyde from stereospecifically deuterated choline

CutC–52 aa from *D. alaskensis* G20 was first activated to the glycyl radical form as described above. After 1 hr of incubation, this mixture was diluted 15-fold. Each coupled assay contained 25 mM Tris HCl buffer pH 8, 50 mM NaCl, 80 mM NADH, 20 μM yeast alcohol dehydrogenase, 0.5 μM CutC dimer–52 aa, 2 μM CutD, 0.2 mM NaDT, 0.2 mM SAM, and substrate (50 mM unlabeled choline, 38.9 mM (*R*)-[1-^{2}H]-choline, or 50 mM (*S*)-[1-^{2}H]-choline) in a total volume of 600 μL. The assays were incubated in an Mbraun anaerobic chamber for 1 h before quenching with 15.4 μL formic acid and centrifuging at 13,000 rpm for 10 min to remove precipitated protein. Each resulting supernatant was transferred to an NMR tube together with a small amount of DMSO-d_6 (final concentration of 5 mM), and analyzed by ^{2}H NMR as described in 'materials and general methods'.

3.3.20: Construction of the CutC homology model

The work described in this section and in 3.3.20 was conducted by Jonathan Marks. The Basic Local Alignment Search Tool (BLAST) (51) was used to search the protein databank (PDB) to find templates for homology model creation, and templates were selected to maximize sequence identity and query coverage. Potential template structures with the highest identity were all GREs: glycerol dehydratase (GDH) from *Clostridium butyricum* (PDB ID: 1R8W) (13), a putative GDH from *Archaeoglobus fulgidus* (PDB ID: 2F3O) (52), a Kolbe-type decarboxylase from *Clostridium scatologenes* (PDB ID: 2Y8N) (53), and pyruvate formate lyase (PFL) from *E. coli* (PDB ID: 2PFL) (54). Modeller v9.12 was used to generate five randomly-seeded homology models of CutC with each

individual potential template structure and with multiple alignment of the top two, top three, and top four template structures. Heteroatoms and water molecules were excluded from the calculations and multiple template alignment was performed using Salign (55). Following model generation using the automodel algorithm, models were optimized using Modeller's built-in conjugate gradient optimization algorithm set to the highest setting and optimization was repeated twice per model. Models with the lowest Discrete Optimization Potential Energy (DOPE) score from each set of homology models were compared based on their DOPE profiles and putative active site geometries to determine the best set of template structures for homology model creation. The homology model based on the multiple alignment of GDH (1R8W) and the putative GDH GRE from *A. fulgidus* (2F3O) was selected due to its low DOPE score and favorable geometry, and 100 randomly-seeded models were generated using the same algorithm as above. The homology model with the lowest DOPE score was analyzed with the verify3D (56), PROCHECK (57), QMEAN6 (58), ERRAT (59), WHATCHECK (60), and PROVE (61) packages.

The CutC homology model is consistent with known GRE structures and passes multiple structure quality benchmarks (**Table 3.2**). The canonical GRE fold with Gly821 and Cys489 surrounded by the characteristic 10-stranded α/β barrel within the homology model aligns well with other members of the GRE family. With an overall DOPE score that is within 3.1-5.5% of the published GDH and putative GDH (*A. fulgidus*) structures, respectively, and with a QMean score that is at the 55th percentile of all known structures (Z-score of 0.142), the CutC homology model is compatible with experimentally-determined structures. In the DOPE profile (**Figure 3.18.A**) and the 3D-1D profile of the model (**Figure 3.18.B**), there are few notable deviations from the template structure profiles, with the exception of the increased DOPE scores for the 209-220 loop in the predicted CutC structure. The Ramachandran plot statistics for the homology model reveal that 99.2% of the model residues are within allowable regions and fewer than 0.1% occupy disallowed regions, indicating good stereochemical quality (**Figure 3.18.C**).

3.3.21: Induced fit docking of choline into the active site of the CutC homology model

The Glide (30, 62) and Prime (63) programs within the Schrödinger suite of molecular dynamics programs have been successfully used in the past to detect and model enzyme active sites and substrate

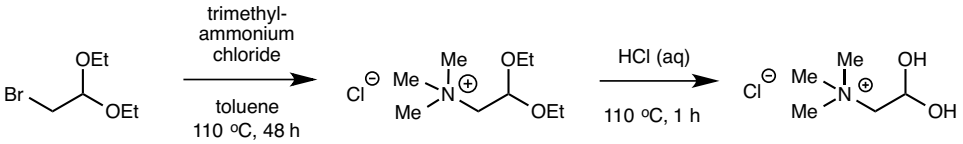
binding orientations (64). In this study, Schrödinger Suite 2012 was used for active site refinement of CutC and induced-fit docking of calculated conformations of the choline substrate.

The standard protein preprocessing protocol of assigning bond orders, adding hydrogen atoms, and capping termini was followed to prepare the selected receptor input file. Hydrogen bond assignment was optimized using PROPKA with a pH of 7.4, and a restrained minimization using the OPLS 2005 force field was used to resolve a single atom-atom clash produced by hydrogen addition. The choline ligand was prepared using the LigPrep algorithm with an approximate pH of 7.4, and ligand minimizations were carried out using the OPLS 2005 force field.

Preliminary docking was performed with Glide using SP precision and van der Waals scaling of 0.5. The receptor grid (18.65Å x 18.65Å x 18.65Å) was centered on the centroid of the active site residues Asp216, Thr334, Phe395, Cys489, Glu491, and Thr502. Following preliminary docking, Prime was used to optimize the active site. All amino acids within 5.0 Å of the docked choline were refined and all putative active site residues were marked for additional refinement. Glide docking with XP precision was then used to re-dock the choline ligand within the refined active site, and calculated poses were compared based on their Glide XP docking scores, Glide Emodel scores, and Prime Energy scores as well as active site geometry. The scores for the best model were Glide: –4.95, Emodel: –38.55, and Prime: 54491.94.

3.3.22: Chemical synthesis procedures and characterization data

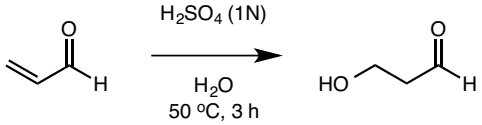
Synthesis of 2,2-dihydroxy-*N,N,N*-trimethylethan-1-aminium chloride (betaine aldehyde)



Betaine aldehyde (2,2-dihydroxy-*N*,*N*,*N*-trimethylethan-1-aminium chloride) was synthesized following a previously published procedure (65), with the exception that the recrystallization step was replaced with an ethyl acetate extraction to remove any unreacted acetal. Concentration *in vacuo* of the HCl-catalyzed hydrolysis reaction afforded 2,2-dihydroxy-*N*,*N*,*N*-trimethylethan-1-aminium chloride as a

light yellow solid (0.389 g, 49%). ^{1}H NMR: (500 MHz, D_2O) 5.55 (t, 1H, C**H**$(OH)_2$, J = 5.5 Hz), 3.42 (d, 2H, NC**H**$_2$, J = 5.5 Hz), 3.22 (s, 9H, N(C**H**$_3)_3$). ^{13}C NMR (100 MHz, DMSO-d_6): 86.28, 69.86, 55.50. HRMS (ESI): calcd for $C_5H_{12}NO^+$ $[M]^+$, 102.0913; found, 102.1071.

Synthesis of 3-hydroxypropanal



3-Hydroxypropanal was synthesized following a published procedure (66), with the following modification: upon neutralization with calcium carbonate, the final mixture was filtered and stored at 4 ºC. The identity of the resulting 3-hydroxypropanal product was confirmed through GC-MS/MS. The ^{13}C NMR data (**Figure 3.28**) also matched that reported previously, revealing a mixture of 3-HPA and 2-(2-hydroxyethyl)-4-hydroxy-1,3-dioxane (HPA dimer) (66).

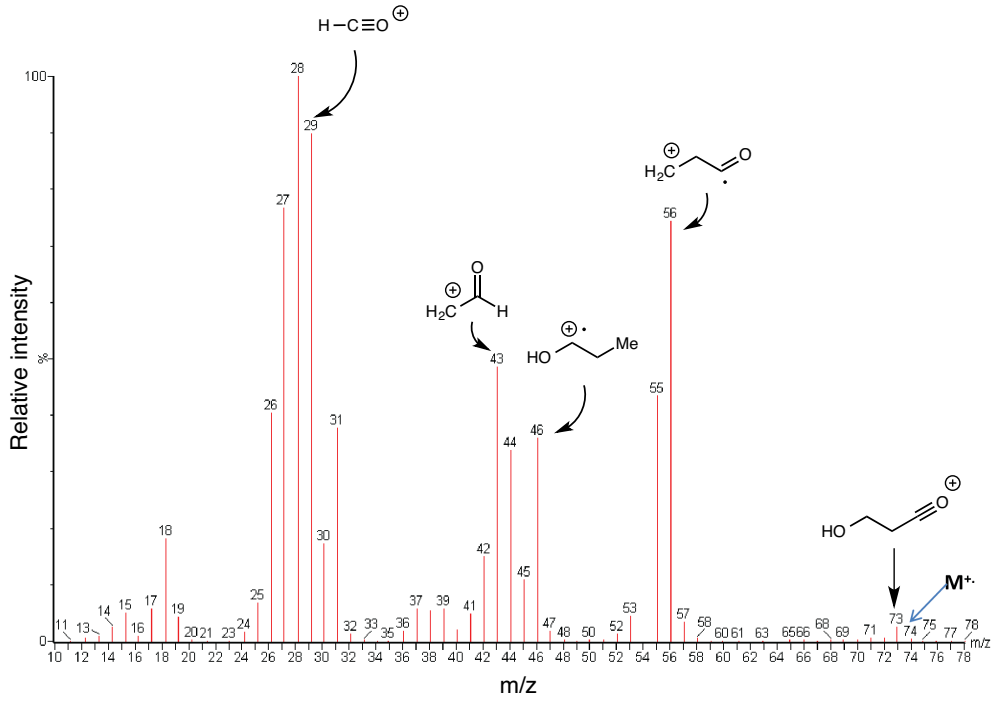


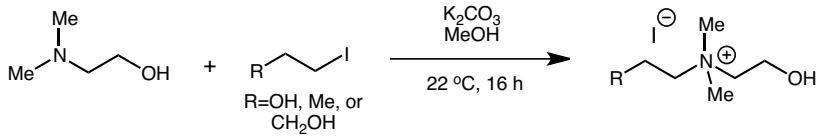
Figure 3.28: GC-MS/MS spectrum of synthesized 3-hydroxypropanal.

Synthesis of (*R*)-1-hydroxy-*N,N,N*-trimethylpropan-2-aminium ((*R*)-2-methylcholine)

To a solution of (*R*)-2-amino-1-propanol (43 mg, 0.57 mmol) in methanol (3.6 mL), potassium carbonate (306 mg, 2.22 mmol) and 1-iodomethane (409 mg, 0.18 mL, 2.88 mmol) were added, and the

mixture was stirred overnight at room temperature, under a nitrogen atmosphere. The mixture was filtered, and the solvent removed *in vacuo*. Synthesis of the (*S*) enantiomer was performed in the same manner, except that (*S*)-2-amino-1-propanol was used as starting material. The weight of the resulting white solids was higher than the theoretical yield, likely due to the presence of potassium iodide since no organic impurities seemed present by NMR. Thus, the yield of each reaction was approximated by ^{1}H NMR, using a known amount of ethyl acetate as standard (38% yield for (*R*)-2-methylcholine; 65% yield for (*S*)-2-methylcholine). HRMS (ESI): calc'd for $C_6H_{16}NO^+$ $[M]^+$, 118.1232; found, 118.1231. ^{1}H-NMR (500 MHz; CD_3OD) δ: 4.41 (d, *J* = 5.3 Hz, 1H, N-C**H**), 3.48 (dd, *J* = 21.2, 13.5 Hz, 1H, C**H**-OH), 3.37 (dd, *J* = 12.7, 11.0 Hz, 1H, C**H**-OH), 3.29 (s, 9H, N-C**H**$_3$), 1.27 (d, *J* = 6.2 Hz, 3H, CH-C**H**$_3$). ^{13}C NMR (126 MHz; CD_3OD) δ: 72.6, 63.4, 55.1, 39.1, 22.29, 22.25.

Synthesis of *N*-(2-hydroxyethyl)-*N*,*N*-dimethylpropan-1-aminium iodide



To a solution of 2-dimethylaminoethanol (200 mg, 2.03 mmol) in methanol (MeOH) (5 mL), potassium carbonate (364 mg, 2.64 mmol) and 1-iodopropane (690 mg, 396 µL, 4.06 mmol) were added, and the mixture was stirred overnight at room temperature, under a nitrogen atmosphere. The mixture was filtered, and the solvent was removed *in vacuo*. The resulting solid was dissolved in water and washed twice with diethyl ether and twice with ethyl acetate. The water was removed *in vacuo* to afford a white solid (333 mg, 1.29 mmol, 63% yield). HRMS (ESI): calc'd for $C_7H_{18}NO^+$ $[M]^+$, 132.1388; found, 132.1393. ^{1}H-NMR (500 MHz; CD_3OD) δ: 4.01 (m, 2H, C**H**$_2$OH), 3.48 (t, *J* = 4.9 Hz, 2H, C**H**$_2$CH$_2$OH), 3.37 (dt, *J* = 8.2, 4.3 Hz, 2H, C**H**$_2$N), 3.16 (s, 6H, C**H**$_3$N), 1.87-1.79 (m, 2H, C**H**$_2$-CH$_3$), 1.01 (t, *J* = 7.3 Hz, 3H, C**H**$_3$CH$_2$). ^{13}C NMR (126 MHz; CD_3OD) δ: 68.1, 66.6, 56.7, 52.41, 52.35, 17.1, 10.9.

Synthesis of 3-hydroxy-*N*-(2-hydroxyethyl)-*N*,*N*-dimethylpropan-1-aminium iodide

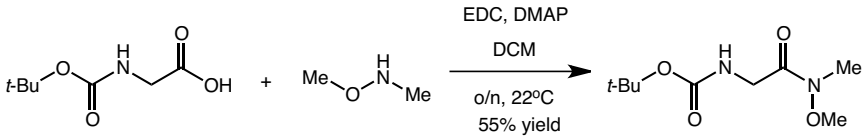
The same procedure was employed as for *N*-(2-hydroxyethyl)-*N*,*N*-dimethylpropan-1-aminium iodide, except that 3-iodo-1-propanol (755 mg, 0.39 mL, 4.06 mmol) was used instead of 1-iodopropane.

The product was obtained as a white solid (521.5 mg, 1.89 mmol, 93 % yield). HRMS (ESI): calc'd for $C_7H_{18}NO_2^+$ $[M]^+$, 148.1338; found, 148.134. ^{1}H-NMR (500 MHz; CD_3OD) δ: 4.08 (m, 2H, C**H**$_2$-OH), 3.72 (t, *J* = 5.8 Hz, 2H, N-C**H**$_2$), 3.65 (m, 4H, N-C**H**$_2$, C**H**$_2$-OH), 3.31 (s, 6H, N-C**H**$_3$), 2.12-2.06 (m, 2H, C**H**$_2$-CH$_2$-OH). ^{13}C NMR (126 MHz; CD_3OD) δ: 66.5, 64.2, 59.2, 56.7, 52.77, 52.74, 26.8.

Synthesis of *N*-(2-hydroxyethyl)-*N*,*N*-dimethylbutan-1-aminium iodide

The same procedure was employed as for *N*-(2-hydroxyethyl)-*N*,*N*-dimethylpropan-1-aminium iodide, except that 1-iodobutane (747 mg, 0.46 mL, 4.06 mmol) was used instead of 1-iodopropane. The product was obtained as a white solid (564 mg, 2.02 mmol, 99 % yield). HRMS (ESI): calc'd for $C_8H_{20}NO^+$ $[M]^+$, 146.1545; found, 146.1551. ^{1}H-NMR (500 MHz; CD_3OD) δ: 4.08 (d, *J* = 1.8 Hz, 2H, C**H**$_2$-OH), 3.65 (t, *J* = 4.6 Hz, 2H, N-C**H**$_2$), 3.59-3.56 (m, 2H, N-C**H**$_2$), 1.86 (dt, *J* = 15.8, 7.9 Hz, 2H, CH$_2$-C**H**$_2$-CH$_2$), 1.47 (q, *J* = 7.4 Hz, 2H, C**H**$_2$-CH$_3$), 1.05 (t, *J* = 7.4 Hz, 3H, CH$_2$-C**H**$_3$). ^{13}C NMR (126 MHz; CD_3OD) δ: 66.37, 66.29, 56.7, 52.6, 48.8, 25.5, 20.4, 14.0.

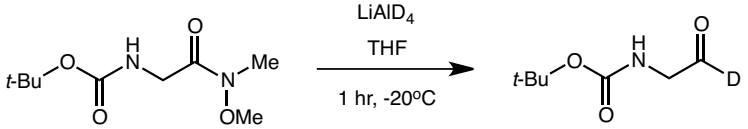
Synthesis of *tert*-butyl (2-(methoxy(methyl)amino)-2-oxoethyl)carbamate



To a solution of N-(*tert*-butoxycarbonyl)-glycine (1.05 g, 6 mmol) in dichloromethane (DCM) (30 mL), *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (EDC) (1.4 g, 9 mmol), 4-(dimethylamino)pyridine (DMAP) (1.1 g, 9 mmol), and N,O-dimethylhydroxylamine hydrochloride (0.55 g, 9 mmol) were added, and the mixture was stirred overnight at room temperature, under a nitrogen atmosphere. The reaction was quenched by the addition of brine (30 mL), the layers were separated, and the aqueous layer was extracted with ethyl acetate (3 x 20 mL). The combined organic layers were washed with 1 M hydrochloric acid (2 x 20 mL) and with brine (20 mL), then dried over magnesium sulfate, filtered and concentrated *in vacuo*. The resulting white solid was purified with flash chromatography, eluting with 80% ethyl acetate in hexanes to give the pure product as a white solid (0.72 g, 3.3 mmol, 55% yield). TLC: Rf=0.4 (silica gel, 1:1 ethyl acetate:hexanes). HRMS (ESI): calc'd for

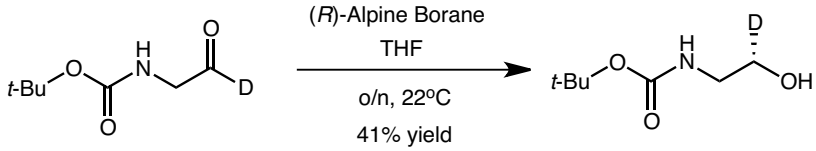
$C_9H_{19}N_2O_4^+$ $[M+Na]^+$, 241.1159; found, 241.1159. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 5.26 (s, 1H, N**H**), 4.03 (s, 2H, C**H**$_2$-NH), 3.67 (s, 3H, O-C**H**$_3$), 3.16 (s, 3H, N-C**H**$_3$), 1.41 (s, 9H, C-C**H**$_3$). ^{13}C NMR (126 MHz; $CDCl_3$) δ: 169.5, 156.0, 79.7, 61.6, 41.8, 32.5, 28.60, 28.56, 28.3.

Synthesis of *tert*-butyl (2-oxo-[2-^{2}H]-ethyl)carbamate



To a solution of *tert*-butyl (2-(methoxy(methyl)amino)-2-oxoethyl)carbamate (360 mg, 1.65 mmol) in tetrahydrofuran (THF) (9.5 mL) cooled to –20 ºC (using a brine-ice bath), a solution of lithium aluminum deuteride in THF (1 M, 3.3 mL, 138.3 mg, 3.3 mmol) was added dropwise, under an argon atmosphere. The reaction mixture was stirred for 1 h then diluted with cold diethyl ether (10 mL), and quenched by the dropwise addition of citric acid (20 % w/v, 0.5 mL). The reaction mixture was further diluted with citric acid (20 % w/v, 10 mL), and the aqueous layer was extracted with diethyl ether (3 x 10 mL). The organic layers were combined, dried over magnesium sulfate and concentrated *in vacuo*. Care was taken to remove most of the solvent as quickly as possible and the product was immediately used in the following reaction without any purification, due to its instability. TLC: Rf=0.3 (silica gel, 1:1 ethyl acetate:hexanes). ^{1}H-NMR (500 MHz; $CDCl_3$) δ: 5.21 (s, 1H, N**H**), 4.07 (d, J = 5.0 Hz, 2H, C**H**$_2$-NH), 1.44 (s, 9H, C-C**H**$_3$).

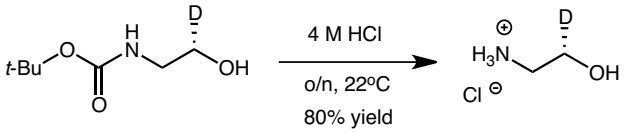
Synthesis of (*S*)-*tert*-butyl ([2-^{2}H]-2-hydroxy-ethyl)carbamate



To a solution of *tert*-butyl (2-oxo-[2-^{2}H]-ethyl)carbamate (230 mg, 1.44 mmol) in THF (8.6 mL) cooled to 0 ºC, a solution of (*R*)-Alpine Borane in THF (0.5 M, 5.74 mL, 2.87 mmol) was added dropwise under an argon atmosphere. The ice bath was removed and the reaction mixture was stirred at room temperature for overnight (o/n). Upon cooling to 0 ºC, the excess Alpine Borane was quenched with ~50

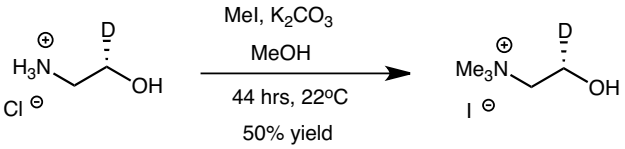
μL acetaldehyde, while stirring under an argon atmosphere for 1 h. The reaction mixture was concentrated *in vacuo*, then diluted with diethyl ether and cooled to 0 ºC. A solution of ethanolamine (262 μL, 266 mg, 4.35 mmol) in THF (0.5 mL) was added slowly to the reaction, the ice bath was removed, and the mixture was stirred at room temperature for 2 h. The white precipitate that formed during this time was dissolved in water, and the mixture was transferred to a separatory funnel. The organic layer was washed with 0.1 M HCl (10 mL), then brine (2 x 10 mL), dried over magnesium sulfate, filtered and concentrated *in vacuo*. The residue was purified by flash chromatography eluting with 80-100% ethyl acetate in hexanes to give the pure product as yellow oil (95 mg, 0.59 mmol, 41% yield). TLC: Rf=0.25 (silica gel, 1:1 ethyl acetate:hexanes). The synthesis of (*R*)-tert-butyl ([2-^{2}H]-2-hydroxy-ethyl)carbamate was carried out in the same manner, with the exception that (*S*)-Alpine Borane was used as reducing agent. HRMS (ESI): calc'd for $C_7H_{15}DNO_3^+$ $[M+Na]^+$, 185.1007; found, 185.1008. ^{1}H-NMR (500 MHz; CD_3OD) δ: δ 3.53 (t, *J* = 5.5 Hz, 1H, OH-C**H**D), 3.14 (d, *J* = 5.7 Hz, 2H, C**H**$_2$-NH), 1.44 (s, 9H, C-C**H**$_3$). ^{13}C NMR (126 MHz; CD_3OD) δ: 130.5, 80.0, 61.83, 61.66, 61.49, 43.7, 28.7.

Synthesis of (*S*)-[1-^{2}H]-ethanolamine hydrochloride



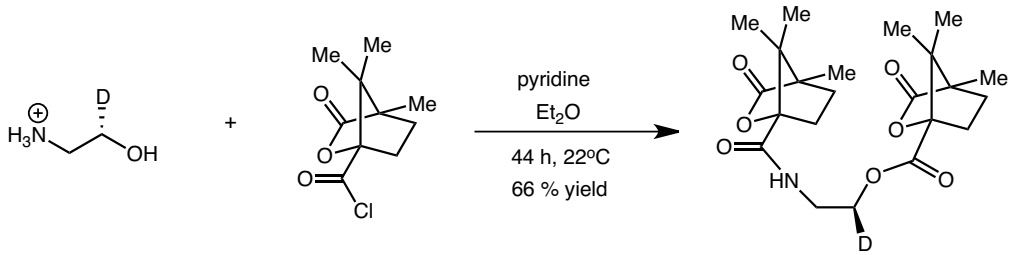
A solution of (*S*)-*tert*-butyl ([2-^{2}H]-2-hydroxy-ethyl)carbamate (70 mg, 0.43 mmol) in hydrochloric acid (4 M, 3 mL) was stirred overnight at room temperature. The reaction mixture was transferred to a separatory funnel and washed with diethyl ether (3 x 10 mL). The water and hydrochloric acid were removed *in vacuo* to afford an off-white solid (34 mg, 0.35 mmol, 80% yield). Note that any impurities present in the starting material will become a brown-black polymer after this reaction. The synthesis of (*R*)-[1-^{2}H]-ethanolamine hydrochloride was carried out in the same manner from the (*R*) enantiomer of the starting material. HRMS (ESI): calc'd for C_2H_6DNO $[M+Na]^+$, 85.0483; found, 84.9595. ^{1}H-NMR (500 MHz; CD_3OD) δ: 3.74 (m, 1H, C**H**-OH), 3.03 (d, *J* = 4.6 Hz, 2H, C**H**$_2$-NH_3). ^{13}C NMR (126 MHz; CD_3OD) δ: 58.70, 58.52, 58.35, 42.7.

Synthesis of (*S*)-[1-^{2}H]-choline iodide



To a solution of (*S*)-[1-^{2}H]-ethanolamine hydrochloride (24.5 mg, 0.25 mmol) in methanol (2.4 mL), potassium carbonate (132 mg, 0.96 mmol) and 1-iodomethane (176 mg, 77 µL, 1.24 mmol) were added, and the mixture was stirred overnight at room temperature, under a nitrogen atmosphere. The mixture was filtered, and the solvent removed *in vacuo*. Synthesis of the (*S*) enantiomer was performed in the same manner, except that (*S*)-[1-^{2}H]-ethanolamine hydrochloride was used as starting material. The weight of the resulting white solids was higher than the theoretical yield, likely due to the presence of potassium iodide since no organic impurities seemed present by NMR. Thus, the purity of each product was approximated by LC-MS/MS, using (trimethyl-d_9)-choline as internal standard. (*S*)-[1-^{2}H]-choline is 13.9% pure, 93% monodeuterated, 4.4% dideuterated, 2.6% unlabeled. (*R*)-[1-^{2}H]-choline is 10.8% pure, 83% monodeuterated, 13% dideuterated, 4% unlabeled. HRMS (ESI): calc'd for $C_5H_{13}DNO^+$ $[M]^+$, 105.1138; found, 105.1140. ^{1}H-NMR (500 MHz; CD_3OD) δ: 3.99 (s, 1H, C**H**-OH), 3.52 (d, J = 4.8 Hz, 2H, N-C**H**$_2$), 3.24 (s, 9H, N-C**H**$_3$). ^{13}C NMR (126 MHz; CD_3OD) δ: 69.0, 54.85, 54.82, 54.79, 52.2.

Synthesis of the N,O-dicamphanoyl derivative of (*S*)-[1-^{2}H]-ethanolamine hydrochloride



To a suspension of (*S*)-[1-^{2}H]-ethanolamine hydrochloride (12 mg, 0.12 mmol) in diethyl ether (1.3 mL), dry pyridine (511 mg, 523 µL, 6.5 mmol) and (1*S*)-(–)-camphanic chloride (130 mg, 0.6 mmol) were added, and the mixture was stirred for 44 h under a nitrogen atmosphere. The excess (–)-camphanic chloride was quenched by the addition of 2 mL of water, and the mixture was concentrated *in vacuo*. The resulting residue was dissolved in chloroform (5 mL), washed twice with HCl (1 M, 5 mL), saturated

sodium bicarbonate solution (5 mL), and brine (5 mL), dried over magnesium sulfate and concentrated *in vacuo* to afford a white solid (33.7 mg, 0.08 mmol, 66 % yield). The syntheses of the (*R*) enantiomer and the unlabeled derivative were carried out in the same manner, using (*R*)-[1-^{2}H]-ethanolamine hydrochloride or ethanolamine hydrochloride as starting materials, respectively. HRMS (ESI): calc'd for $C_{22}H_{31}DNO$ $[M+Na]^{+}$, 445.2056; found, 445.2064. ^{13}C NMR (126 MHz; CD_3OD) δ: 180.08, 180.00, 169.6, 168.7, 111.4, 93.7, 92.7, 56.4, 56.0, 55.3, 54.9, 39.1, 31.6, 31.2, 30.4, 29.95, 29.89, 17.13, 17.05, 17.02, 9.93, 9.87. ^{1}H-NMR for the (*S*) enantiomer (500 MHz; $CDCl_3$) δ: 6.76 (t, *J* = 5.6 Hz, 1H), 4.31 (dd, *J* = 6.6, 4.2 Hz, 1H), 3.66 (ddd, *J* = 14.5, 6.0, 4.2 Hz, 1H), 3.57 (dt, *J* = 14.5, 6.4 Hz, 1H), 2.48 (ddd, *J* = 13.2, 10.4, 4.2 Hz, 1H), 2.39 (ddd, *J* = 13.5, 10.8, 4.2 Hz, 1H), 2.01 (ddd, *J* = 13.6, 9.3, 4.5 Hz, 1H), 1.94-1.83 (m, 3H), 1.66 (ddd, *J* = 13.4, 9.1, 4.5 Hz, 2H), 1.08 (s, 3H), 1.07 (s, 3H), 1.07 (s, 3H), 1.03 (s, 3H), 0.93 (s, 3H), 0.88 (s, 3H). ^{1}H-NMR for the (*R*) enantiomer (500 MHz; $CDCl_3$) δ: 6.73 (t, *J* = 5.5 Hz, 1H, N**H**), 4.30 (dd, *J* = 6.5, 4.2 Hz, 1H, O-C**H**D), 3.70-3.66 (m, 1H, CD-C**H**), 3.59 (ddd, *J* = 14.6, 6.0, 4.3 Hz, 1H, CD-C**H**), 2.50 (ddd, *J* = 13.1, 10.3, 4.2 Hz, 1H), 2.41 (ddd, *J* = 13.5, 10.8, 4.2 Hz, 1H), 2.03 (ddd, *J* = 13.6, 9.3, 4.5 Hz, 1H), 1.96-1.84 (m, 3H), 1.67 (dtd, *J* = 13.7, 9.3, 4.9 Hz, 2H), 1.10 (s, 3H), 1.09 (s, 3H), 1.09 (s, 3H) 1.05 (s, 3H), 0.96 (s, 3H), 0.89 (s, 3H). ^{1}H-NMR for the unlabeled product (500 MHz; $CDCl_3$) δ: 6.78 (t, *J* = 5.6 Hz, 1H), 4.34-4.26 (m, 2H), 3.65 (dtd, *J* = 14.5, 6.2, 4.5 Hz, 1H), 3.57 (dtd, *J* = 14.5, 6.2, 4.5 Hz, 1H), 2.47 (ddd, *J* = 13.2, 10.5, 4.2 Hz, 1H), 2.38 (ddd, *J* = 13.5, 10.8, 4.2 Hz, 1H), 2.00 (ddd, *J* = 13.6, 9.2, 4.5 Hz, 1H), 1.93-1.81 (m, 3H), 1.64 (ddt, *J* = 13.4, 9.0, 4.5 Hz, 2H), 1.07 (s, 3H), 1.06 (s, 3H), 1.06 (d, *J* = 1.2 Hz, 3H), 1.01 (s, 3H), 0.91 (s, 3H), 0.86 (s, 3H).

3.4: References

1. Selmer T, Pierik AJ, & Heider J (2005) New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem* 386(10):981-988.

2. Craciun S & Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109(52):21307-21312.

3. Hayward HR & Stadtman TC (1960) Anaerobic degradation of choline. II. Preparation and properties of cell-free extracts of *Vibrio cholinicus*. *J Biol Chem* 235:538–543.

4. Sandhu SS & Chase T (1986) Aerobic degradation of choline by *Proteus mirabilis*: enzymatic requirements and pathway. *Can J Microbiol* 32(9):743-750.

5. Zeisel SH & da Costa KA (2009) Choline: an essential nutrient for public health. *Nutr Rev* 67(11):615-623.

6. Dumas ME*, et al.* (2006) Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci USA* 103(33):12511-12516.

7. Wang Z*, et al.* (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57-63.

8. Christodoulou J (2012) Trimethylaminuria: An under-recognised and socially debilitating metabolic disorder. *J Paediatr Child Health* 48(3):E153-E155.

9. Hippe H, Caspari D, Fiebig K, & Gottschalk G (1979) Utilization of trimethylamine and other *N*-methyl compounds for growth and methane formation by *Methanosarcina barkeri*. *Proc Natl Acad Sci USA* 76(1):494-498.

10. Vey JL*, et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.

11. Fan CG*, et al.* (2010) Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci USA* 107(16):7509-7514.

12. Akita K*, et al.* (2010) Purification and some properties of wild-type and N-terminal-truncated ethanolamine ammonia-lyase of *Escherichia coli*. *J Biochem* 147(1):83-93.

13. O'Brien JR*, et al.* (2004) Insight into the mechanism of the B_{12}-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* 43(16):4635-4645.

14. Unkrig V, Neugebauer FA, & Knappe J (1989) The free radical of pyruvate formate-lyase. Characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analog hypophosphite. *Eur J Biochem* 184(3):723-728.

15. Sun XY*, et al.* (1996) The free radical of the anaerobic ribonucleotide reductase from *Escherichia coli* is at glycine 681. *J Biol Chem* 271(12):6827-6831.

16. Tokumoto U & Takahashi Y (2001) Genetic analysis of the *isc* operon in *Escherichia coli* involved in the biogenesis of cellular iron-sulfur protein. *J Biochem* 130(1):63-71.

17. Hanzelmann P*, et al.* (2004) Characterization of MOCS1A, an oxygen-sensitive iron-sulfur protein involved in human molybdenum cofactor biosynthesis. *J Biol Chem* 279(33):34721-34732.

18. Selvaraj B, Pierik AJ, Bill E, & Martins BM (2013) 4-Hydroxyphenylacetate decarboxylase activating enzyme catalyses a classical S-adenosylmethionine reductive cleavage reaction. *J Biol Inorg Chem* 18(6):633-643.

19. Duee ED*, et al.* (1994) Refined crystal-structure of the 2[4Fe-4] Ferredoxin from *Clostridium acidurici* at 1.84 Å resolution. *J Mol Biol* 243(4):683-695.

20. Lanz ND & Booker SJ (2015) Auxiliary iron-sulfur cofactors in radical SAM enzymes. *Biochim Biophys Acta* 1853(6):1316-1334.

21. Wecksler SR*, et al.* (2009) Pyrroloquinoline quinone biogenesis: demonstration that PqqE from *Klebsiella pneumoniae* is a radical S-adenosyl-L-methionine enzyme. *Biochemistry* 48(42):10151-10161.

22. Yan F*, et al.* (2010) RImN and Cfr are radical SAM enzymes involved in methylation of ribosomal RNA. *J Am Chem Soc* 132(11):3953-3964.

23. Shisler KA & Broderick JB (2014) Glycyl radical activating enzymes: structure, mechanism, and substrate interactions. *Arch Biochem Biophys* 546:64-71.

24. Wagner AFV, Frey M, Neugebauer FA, Schafer W, & Knappe J (1992) The free-radical in pyruvate formate-lyase is located on glycine-734. *Proc Natl Acad Sci USA* 89(3):996-1000.

25. Mulliez E, Padovani D, Atta M, Alcouffe C, & Fontecave M (2001) Activation of class III ribonucleotide reductase by flavodoxin: A protein radical-driven electron transfer to the iron-sulfur center. *Biochemistry* 40(12):3730-3736.

26. Krieger CJ, Roseboom W, Albracht SPJ, & Spormann AM (2001) A stable organic free radical in anaerobic benzylsuccinate synthase of *Azoarcus sp* strain T. *J Biol Chem* 276(16):12924-12927.

27. Yu L, Blaser M, Andrei PI, Pierik AJ, & Selmer T (2006) 4-Hydroxyphenylacetate decarboxylases: Properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* 45(31):9584-9592.

28. Jorda J, Lopez D, Wheatley NM, & Yeates TO (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci* 22(2):179-195.

29. Eswar N*, et al.* (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5:Unit 5-6.

30. Glide, version 5.8, Schrödinger, LLC, New York, NY, 2012.

31. Pittelkow M, Tschapek B, Smits SHJ, Schmitt L, & Bremer E (2011) The crystal structure of the substrate-binding protein OpuBC from *Bacillus subtilis* in complex with choline. *J Mol Biol* 411(1):53-67.

32. Oswald C*, et al.* (2008) Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from *Sinorhizobium meliloti* in the liganded and unliganded-closed states. *J Biol Chem* 283(47):32848-32859.

33. Kolberg M, Strand KR, Graff P, & Andersson KK (2004) Structure, function, and mechanism of ribonucleotide reductases. *Biochem Biophys Acta* 1699(1-2):1-34.

34. Feliks M & Ullmann GM (2012) Glycerol dehydratation by the B_{12}-independent enzyme may not involve the migration of a hydroxyl group: a computational study. *J Phys Chem B* 116(24):7076-7087.

35. Shibata N*, et al.* (2010) Crystal structures of ethanolamine ammonia-lyase complexed with coenzyme B_{12} analogs and substrates. *J Biol Chem* 285(34):26484-26493.

36. Gani D, Wallis OC, & Young DW (1982) Stereochemistry of the coenzyme B_{12}-mediated rearrangement of 2-aminoethan-1-ol by ethanolamine ammonia-lyase. *J Chem Soc Chem Comm* (16):898-899.

37. Midland MM (1989) Asymmetric reductions with organoborane reagents. *Chem Rev* 89(7):1553-1561.

38. Gani D & Young DW (1982) Synthesis of (2*S*,3*R*)[3-$^2H_1$]- and (2*S*,3*S*)[2,3-$^2H_2$]-serines and (1*R*)[1-$^2H_1$]- and (1*S*,2*RS*)[1,2-$^2H_2$]-2-aminoethanols. *J Chem Soc Chem Comm* (15):867-869.

39. Larkin MA*, et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.

40. Kearse M*, et al.* (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.

41. Stoscheck CM (1990) Quantitation of protein. *Method Enzymol* 182:50-68.

42. Stoll S & Schweiger A (2006) EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J Magn Reson* 178(1):42-55.

43. Murib JH & Ritter DM (1952) Decomposition of nitrosyl disulfonate ion. I. Products and mechanism of color fading in acid solution. *J Am Chem Soc* 74(13):3394-3398.

44. Hall DA, Jordan-Starck TC, Loo RO, Ludwig ML, & Matthews RG (2000) Interaction of flavodoxin with cobalamin-dependent methionine synthase. *Biochemistry* 39(35):10711-10719.

45. Szu PH, Ruszczycky MW, Choi SH, Yan F, & Liu HW (2009) Characterization and mechanistic studies of Desll: a radical S-adenosyl-L-methionine enzyme involved in the biosynthesis of TDP-D-desosamine. *J Am Chem Soc* 131(39):14030-14042.

46. Kennedy MC*, et al.* (1984) Evidence for the formation of a linear [3Fe-4S] cluster in partially unfolded aconitase. *J Biol Chem* 259(23):4463-4471.

47. Beinert H (1983) Semi-micro methods for analysis of labile sulfide and of labile sulfide plus sulfane sulfur in unusually stable iron sulfur proteins. *Anal Biochem* 131(2):373-378.

48. Ollagnier S*, et al.* (1997) Activation of the anaerobic ribonucleotide reductase from *Escherichia coli* - the essential role of the iron-sulfur center for S-adenosylmethionine reduction. *J Biol Chem* 272(39):24216-24223.

49. Kulzer R, Pils T, Kappl R, Huttermann J, & Knappe J (1998) Reconstitution and characterization of the polynuclear iron-sulfur cluster in pyruvate formate-lyase-activating enzyme - molecular properties of the holoenzyme form. *J Biol Chem* 273(9):4897-4903.

50. Crain AV & Broderick JB (2014) Pyruvate formate-lyase and its activation by pyruvate formate-lyase activating enzyme. *J Biol Chem* 289(9):5723-5729.

51. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410.

52. Lehtio L, Grossmann JG, Kokona B, Fairman R, & Goldman A (2006) Crystal structure of a glycyl radical enzyme from *Archaeoglobus fulgidus*. *J Mol Biol* 357(1):221-235.

53. Martins BM*, et al.* (2011) Structural basis for a Kolbe-type decarboxylation catalyzed by a glycyl radical enzyme. *J Am Chem Soc* 133(37):14666-14674.

54. Becker A & Kabsch W (2002) X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA - how the enzyme uses the Cys-418 thiyl radical for pyruvate cleavage. *J Biol Chem* 277(42):40036-40042.

55. Marti-Renom MA, Madhusudhan MS, & Sali A (2004) Alignment of protein sequences by their profiles. *Protein Sci* 13(4):1071-1087.

56. Bowie JU, Luthy R, & Eisenberg D (1991) A method to identify protein sequences that fold into a known 3-dimensional structure. *Science* 253(5016):164-170.

57. Laskowski RA, Macarthur MW, Moss DS, & Thornton JM (1993) Procheck - a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283-291.

58. Benkert P, Biasini M, & Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27(3):343-350.

59. Colovos C & Yeates TO (1993) Verification of protein structures - patterns of nonbonded atomic interactions. *Protein Sci* 2(9):1511-1519.

60. Hooft RWW, Vriend G, Sander C, & Abola EE (1996) Errors in protein structures. *Nature* 381(6580):272-272.

61. Pontius J, Richelle J, & Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol* 264(1):121-136.

62. Halgren TA*, et al.* (2004) Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 47(7):1750-1759.

63. Jacobson MP, Friesner RA, Xiang ZX, & Honig B (2002) On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol* 320(3):597-608.

64. McRobb FM, Capuano B, Crosby IT, Chalmers DK, & Yuriev E (2010) Homology modeling and docking evaluation of aminergic G protein-coupled receptors. *J Chem Inf Model* 50(4):626-637.

65. Landfald B & Strom AR (1986) Choline-glycine betaine pathway confers a high-level of osmotic tolerance in *Escherichia coli*. *J Bacteriol* 165(3):849-855.

66. Vollenweider S, Grassi G, Konig I, & Puhan Z (2003) Purification and structural characterization of 3-hydroxypropionaldehyde and its derivatives. *J Agr Food Chem* 51(11):3287-3293.

Chapter 4: Molecular basis of C–N bond cleavage by the glycyl radical enzyme choline trimethylamine-lyase[a]

4.1: Introduction

As described in Chapter 1, the anaerobic cleavage of choline by microbes to form acetaldehyde and trimethylamine (TMA) impacts both the environment and human health (**Figure 4.1.A**). Within the guts of ruminants and marine sediments, choline-derived TMA is converted to the powerful greenhouse gas methane by archaea (1, 2). TMA has been estimated to account for a significant portion of the methane produced in marine sediments (35-61%) (2), thus playing a major role in the greenhouse effect, as well as global nitrogen and carbon cycles. Although choline is an important nutrient for humans, conversion of choline to TMA by the gut microbiota and its further oxidation to trimethylamine *N*-oxide by liver enzymes (3) is implicated in a number of diseases, including nonalcoholic fatty liver disease (4), atherosclerosis (5, 6) and the metabolic disorder trimethylaminuria (fish malodor syndrome) (7). Chapter 2 of this thesis described the discovery of the choline utilization (*cut*) gene cluster in *Desulfovibrio* species and the validation of its function (8). The key enzyme in this pathway is a novel C–N bond cleaving glycyl radical enzyme (GRE) choline trimethylamine-lyase (CutC), which converts choline into TMA and acetaldehyde (9). Subsequent bioinformatic analyses performed by Balskus group members have revealed that choline utilization is highly prevalent in human stool metagenomes and is widely distributed among both environmental and human-associated bacteria (10), highlighting the importance of CutC for TMA production in diverse anoxic microbial habitats.

[a] This chapter is an unofficial adaptation of an article submitted to *Cell Chemical Biology*, which was written together with Prof. Emily P. Balskus, Prof. Catherine L. Drennan (MIT), and Dr. Michael A. Funk (MIT).

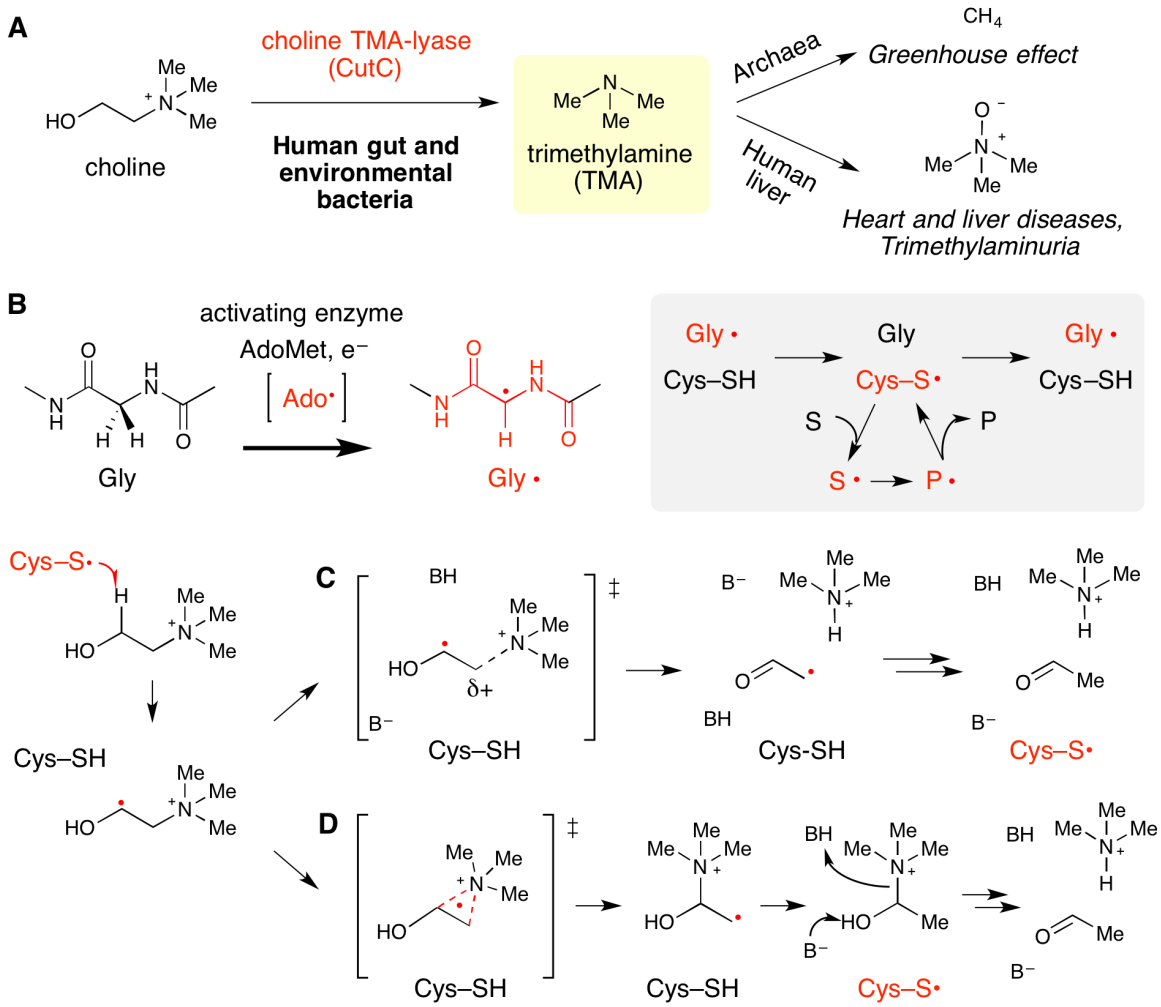


Figure 4.1: Anaerobic metabolism of choline into trimethylamine (TMA) is a disease-associated microbial activity that utilizes the glycyl radical enzyme (GRE) CutC. **(A)** Anaerobic microbes generate TMA from choline in the human gut and in the environment. TMA is further metabolized to the greenhouse gas methane by archaea, or to the disease-associated metabolite trimethylamine *N*-oxide by a human liver monooxygenase. **(B)** General mechanistic hypothesis for GRE function: the 5′-deoxyadenosyl radical (Ado•) is used by an AdoMet radical enzyme to produce a stable glycyl radical within GREs. A thiyl radical is proposed to serve as the active oxidant in all GREs and would be generated transiently within the active site. S = substrate, P = product. Following initial hydrogen atom abstraction to form an α-hydroxy radical, **(C)** CutC may perform base-catalyzed direct elimination of TMA. **(D)** Alternately, choline cleavage could involve a 1,2-migration of the trimethylammonium moiety followed by decomposition of the resulting carbinolamine.

The glycyl radical enzyme (GRE) family has emerged as a remarkable class of biocatalysts that play central roles in anaerobic microbial metabolic pathways, including nucleotide reduction (class III ribonucleotide reductase) and carbon-source utilization (pyruvate formate-lyase) (11). As presented in Chapter 1, all GREs characterized to date must be post-translationally modified by a radical S-adenosylmethionine (AdoMet) activating protein (CutD within the *cut* cluster), which installs a stable, α-

carbon glycyl radical. During GRE catalysis, this glycyl radical species is proposed to generate a transient thiyl radical that initiates the reaction and is regenerated upon product formation (**Figure 4.1.B**). The discovery of CutC expanded the chemistry of GREs to include choline deamination, a net 1,2-elimination reaction that requires a radical enzyme due to the high pKa of the C1 proton of choline. The existence of CutC underscores the diversity of Nature's approaches to radical-mediated C–N bond cleavage, which include two other enzymes: the adenosylcobalamin (AdoCbl)-dependent enzyme ethanolamine ammonia lyase (EAL) (12), and the radical AdoMet-dependent enzyme DesII, involved in desosamine biosynthesis (13). The main differences between CutC and these enzymes are the requirement of the trimethylammonium group for reactivity (9) and the lack of reliance on a 5′-deoxyadenosyl radical (Ado•) for initiating the elimination reaction.

As presented in Chapter 3, we have previously envisioned two general mechanisms for choline cleavage by CutC (9). In a similar manner to other GREs, a CutC thiyl radical could cleave the C1 C–H bond of choline, generating a substrate-based α-hydroxyalkyl radical. Subsequent deprotonation of the alcohol would allow direct elimination of TMA, leading to a resonance stabilized vinoxy radical that could reform the thiyl radical through hydrogen atom abstraction, generating acetaldehyde (**Figure 4.1.C**). This mechanism is analogous to those proposed for the C–O bond cleaving GRE glycerol dehydratase (GDH) (14) and the radical AdoMet-dependent enzyme DesII, which has not yet been structurally characterized (13). Alternatively, the initial substrate based radical could undergo a 1,2-migration of the trimethylammonium group to yield a carbinolamine radical (**Figure 4.1.D**), similar to the generally accepted mechanism for the AdoCbl-dependent enzyme EAL (12). Understanding the structure and mechanism of CutC is important not only for enhancing our knowledge of radical enzymes but also for designing inhibitors of TMA generation. Given that an inhibitor of gut microbial TMA production was found to attenuate atherosclerosis in a mouse model (15), small molecules that target CutC hold promise as tools for studying choline metabolism in complex microbial communities and for further development into therapeutics. Despite the availability of a CutC homology model (described in Chapter 3) and a low-resolution crystal structure (9, 16), many questions remained regarding this enzyme's

interactions with choline and mechanism of catalysis. In this chapter, I present the results of our collaboration with the research group of Prof. Catherine L. Drennan at Massachusetts Institute of Technology. A former graduate student in this group, Dr. Michael A. Funk, elucidated the crystal structures of CutC wild-type and mechanistically informative mutants from *Desulfovibrio alaskensis* G20 in the presence of choline, while I purified these proteins and performed biochemical experiments. Overall, these results provide new insight into the mechanism of this C–N cleaving radical enzyme, thus broadening our understanding of radical-based enzyme catalysis.

4.2: Results and discussion

4.2.1: A conserved GRE architecture is maintained in CutC

The Drennan group crystallized CutC from *D. alaskensis* G20 and solved the structure by molecular replacement with AdoCbl-independent glycerol dehydratase (GDH) (PDB ID 1R8W) (17) as the search model. The four molecules in the asymmetric unit form two dimers with a large shared surface area (**Figure 4.2.A**). This tetrameric packing is similar to that seen in structures of other GREs (**Figure 4.3**); however, size exclusion chromatography experiments indicated previously that CutC is a dimer in solution (9). The assembly observed in the crystal structure may thus be a crystallization artifact or may be relevant within bacterial microcompartments where CutC is thought to be localized (18, 19). Each CutC molecule adopts the canonical GRE fold: a ten-stranded β/α barrel (β1-β10) with two loops containing catalytic residues inserted into the center of the barrel (**Figure 4.2.B**). The Gly loop is located at the tip of the C-terminal glycyl radical domain, a structural unit believed to undergo conformational changes that can close or expose the active site during installation of the glycyl radical by the activating enzyme (20, 21). An absolutely conserved glycine (Gly821 in CutC) is found at the tip of the Gly loop, fully in the interior of the barrel. Next to the Gly loop in the enzyme interior is the Cys loop, which contains an absolutely conserved cysteine residue (Cys489 in CutC) and other residues that contribute to catalysis and are unique to specific GREs (17, 20, 22). The active site cavity is found directly above the Cys loop, consistent with reaction initiation by a transient thiyl radical generated on Cys489 by Gly821.

In agreement with these proposed roles in CutC, C489A and G821A mutants were found to be completely inactive in the production of TMA in an end point assay (9).

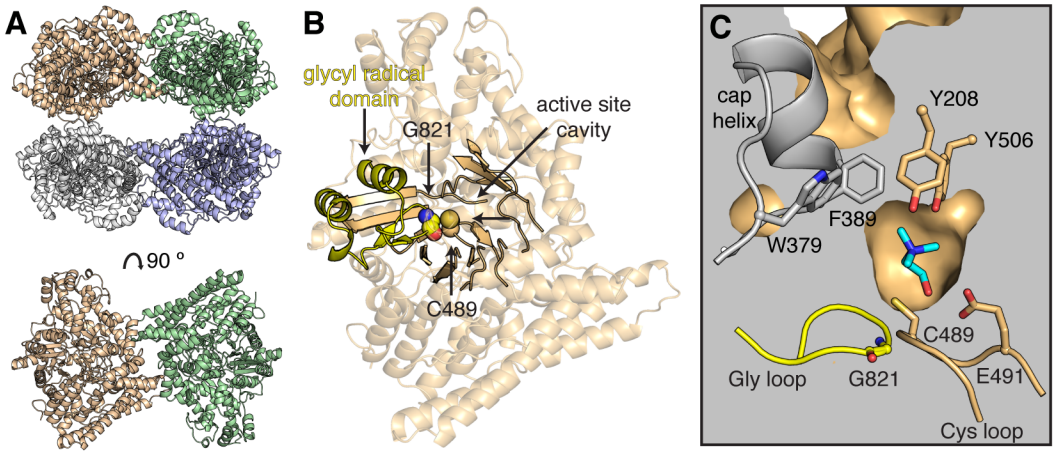


Figure 4.2: Overall structure of CutC from *D. alaskensis* G20. **(A)** CutC crystallizes as a dimer or dimers, similar to several other GREs (see also **Figure 4.3**). The dimeric unit is shown below rotated 90º. **(B)** The active site cavity is found at the center of the $(\beta/\alpha)_{10}$ barrel (solid ribbon), above the Cys loop. Residues C489 (location of putative thiyl radical) and G821 (location of putative glycyl radical) are shown in spheres. The glycyl radical domain (yellow) harbors the active site glycyl radical on the Gly loop in post-translationally modified proteins. **(C)** Cross-section of the CutC active site displaying a ring of aromatic residues (sticks) at the top of the active site, including two residues on a helix that caps the top of the barrel (gray). Choline is shown in the active site (cyan sticks).

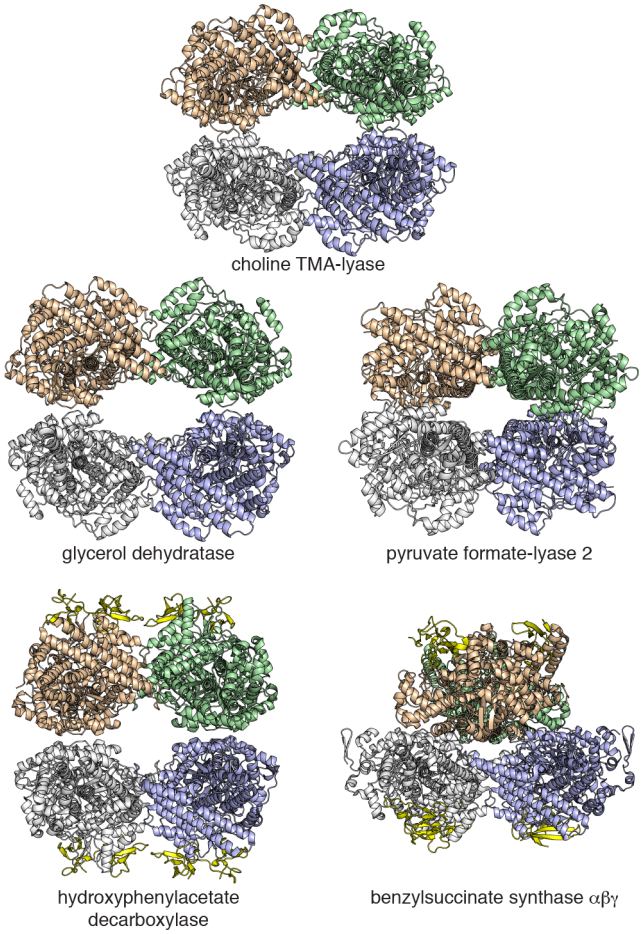


Figure 4.3: Tetrameric assemblies observed in the crystal lattices of GREs. For each GRE, four protomers are shown in roughly the same orientation. Choline TMA-lyase bound to choline, PDB ID 5FAU (this work), is depicted. The other GRE structures shown are PDB ID 1R8W for glycerol dehydratase (GDH) (17), PDB ID 2F3O for pyruvate formate-lyase 2 (PFL2) (23), PDB ID 2YAJ for hydroxyphenylacetate decarboxylase (HPD; small subunit in yellow) (24), and PDB ID 4PKF for benzylsuccinate synthase (BSS; both small subunits in yellow) (25). For the BSSαβγ complex, the upper dimer is oriented 90 º from where it is observed in the other GREs. No tetrameric assembly could be found for the BSSαγ complex (PDB ID 4PKC) or PFL (PDB 1CM5) (20). Only PFL2 and HPD have been observed to adopt a tetrameric (or heterooctameric for HPD) structure in solution (23, 24).

A notable feature of the CutC active site also observed in other GREs (26) is a hydrophobic cap formed by a helix extending from β3 (**Figure 4.2.C**). This cap packs against two tyrosine residues that participate in choline binding (described below) and appears to block access to the active site from the exterior of the protein; simple rotation of the side chains of Phe389 and/or Trp379 into adjacent, solvent-filled regions may allow access to the active site. We hypothesize that the deeply buried active site captured choline during expression because our crystals of wild-type CutC contained a well-ordered

molecule of choline bound, despite dialysis of the protein during purification and no ligand added to the crystallization buffer. Recently, Kalnins et al. solved low resolution choline-bound and choline-free structures of a proteolytic fragment of a homologous CutC from the opportunistic pathogen *Klebsiella pneumoniae* (16). This CutC fragment is of a similar size to the CutC from *D. alaskensis* G20 and utilizes the same active site residues to bind choline. The structures determined by Kalnins et al. complement our work because they reveal that choline can trigger a conformational change in CutC from an open to a closed state and can stabilize the protein towards proteolysis.

4.2.2: Selective binding of choline involves unexpected interactions

The CutC active site is formed by residues from the Cys loop and the top face of the barrel (**Figure 4.4**). Within this Cys loop, the backbone amide nitrogen of Gly488/Cys489 and the carboxylate of Glu491 form hydrogen bonds with the choline hydroxyl group, positioning C1 of choline at a distance suitable (3.6 Å) for pro-*S* hydrogen abstraction by the putative thiyl radical on Cys489. C2 of choline is also in proximity to Cys489 (4.2 Å) potentially facilitating reformation of the thiyl radical following choline cleavage. In turn, the Glu491 carboxylate is positioned by hydrogen bonds from the side chain alcohol of the β6 residue Thr502 and the Cys loop backbone carbonyl. There is also likely a substantial electrostatic component to choline binding that arises from interactions between the diffuse positive charge localized on the trimethylammonium methyl groups (27) and the negatively charged first-shell residues Asp216 and Glu491. The partial positive charge localized on C2 of choline establishes a face-on cation-π interaction with the aromatic ring of β3 residue Phe395 (C2 to ring center: 3.84 Å) (**Figure 4.4.A**). The conservation of the Cys loop sequence motif GCVEP and of an aromatic residue at the Phe395 position in GDH suggests mechanistic similarities between these two classes of GRE lyases (**Figure 4.5**). The similar orientation of choline and glycerol in the active sites of CutC and GDH, respectively, also implies a shared mechanistic logic (17). The choline trimethylammonium group extends away from the Cys loop and is oriented gauche to the hydroxyl, with an average dihedral angle of 61 º in the four molecules in the asymmetric unit.

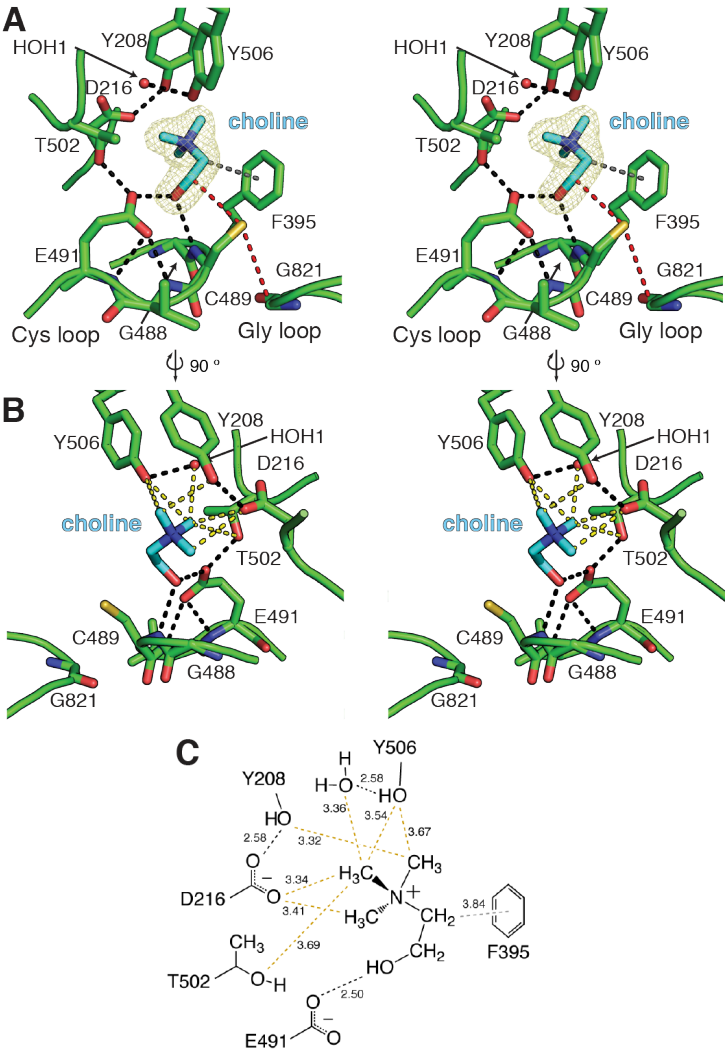


Figure 4.4: CutC binds substrate in a pocket above the Cys loop. **(A)** Stereoimage of CutC bound to its substrate, choline (cyan). The proposed hydrogen atom transfer pathway from G821 to C489 to C1 of choline is marked (red dashes). Hydrogen bonds (2.5-3.2 Å) are shown for residues within the active site (black dashes). The average choline dihedral angle (HO–C1–C2–N(Me_3)) is 61 º. **(B)** Stereoimage of the CH–O bonds (yellow dashes) present in the CutC-choline complex. **(C)** Diagram of protein and water interactions with the trimethylammonium moiety of choline with CH–O and hydrogen bond distances (Å). CH–O bonds are indicated for C to O distances of 3.8 Å or less (yellow). Hydrogen bonds (black) and presumed cation-π interactions (gray) are shown between protein and substrate atoms. Distances (Å) are given in the diagram. The maximum-likelihood-estimated coordinate error is 0.20 Å.

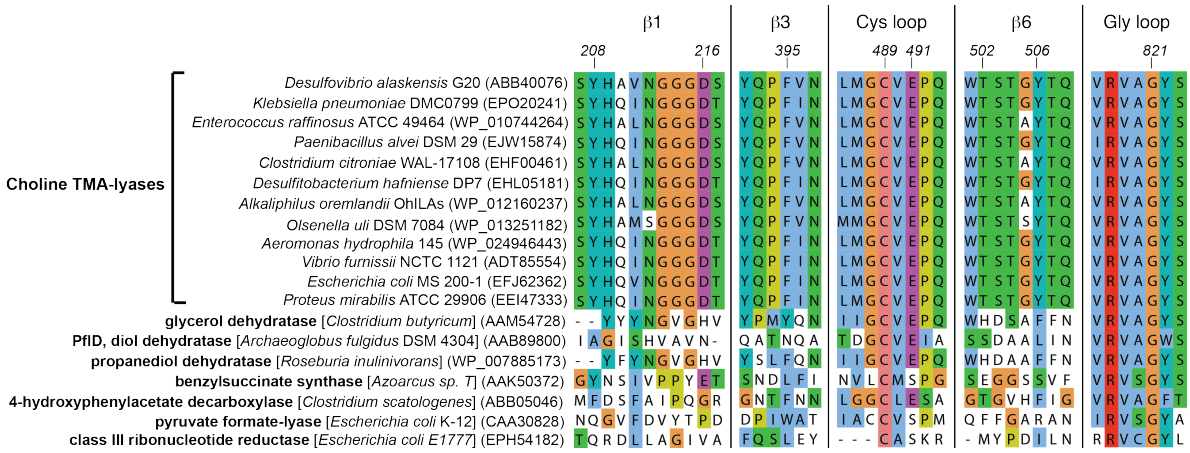


Figure 4.5: Multiple sequence alignment of choline TMA-lyase homologs and characterized glycyl radical enzymes. The full sequence alignment was constructed with Clustal W(28), and regions of interest identified in the CutC structure are shown. The residues are numbered based on the sequence of CutC from *Desulfovibrio alaskensis* G20. Conserved residues are colored according to the residue property (green: polar; blue: nonpolar; purple: acidic; red: basic; orange: Gly; yellow: Pro; pink: Cys). GenBank accession numbers identify the amino acid sequences used.

Unexpectedly, the primary choline-selective interactions observed within the active site involve the trimethylammonium substituent and the side chain oxygen atoms of Tyr208, Tyr506, and Asp216 (**Figure 4.4.B**). These oxygen atoms approach the trimethylammonium methyl groups closely, displaying C to O distances (3.3 and 3.5 Å) shorter than a typical van der Waals distance. Such contacts, previously described as CH–O interactions or bonds, are ubiquitous features in protein and nucleic acid structures and are proposed to play an important, but underappreciated, role in stabilizing macromolecule structure (29). In cases where a partially positive-charged C–H bond is present, CH–O interactions contribute an estimated ~1.2 kcal/mol to the binding energy, comparable to weak hydrogen bonds (30). In the context of SAM-dependent lysine methyltransferases (*e.g.,* SET7/9), these interactions were proposed to perform several functions in lysine methyl transfer, including enhancing AdoMet binding, stabilizing the partially positively charged transition state and aligning the methyl group for optimal transfer geometry (31). Within CutC, there are at least four close CH–O interactions (3.2-3.5 Å) and another three longer interactions (3.5-3.7 Å) (**Figure 4.4.C**). A single water molecule held in the active site by Tyr506 also participates in a close CH–O interaction. Typically, trimethylammonium-binding proteins use aromatic residues to create a hydrophobic pocket and to provide cation-π interactions with the diffuse positive

charge of the trimethylammonium group (27). Comparison of the structures of nine proteins that bind trimethylammonium-containing metabolites (choline, glycine betaine, carnitine, or trimethyllysine) reveals the presence of two to four CH–O interactions in each active site (**Figure 4.6**). Relative to the subset of these proteins that exclusively bind choline (**Figure 4.6.A**), CutC contains a large number of CH–O interactions (seven with distances less than 3.7 Å), which is perhaps reflective of this enzyme's unique reactivity.

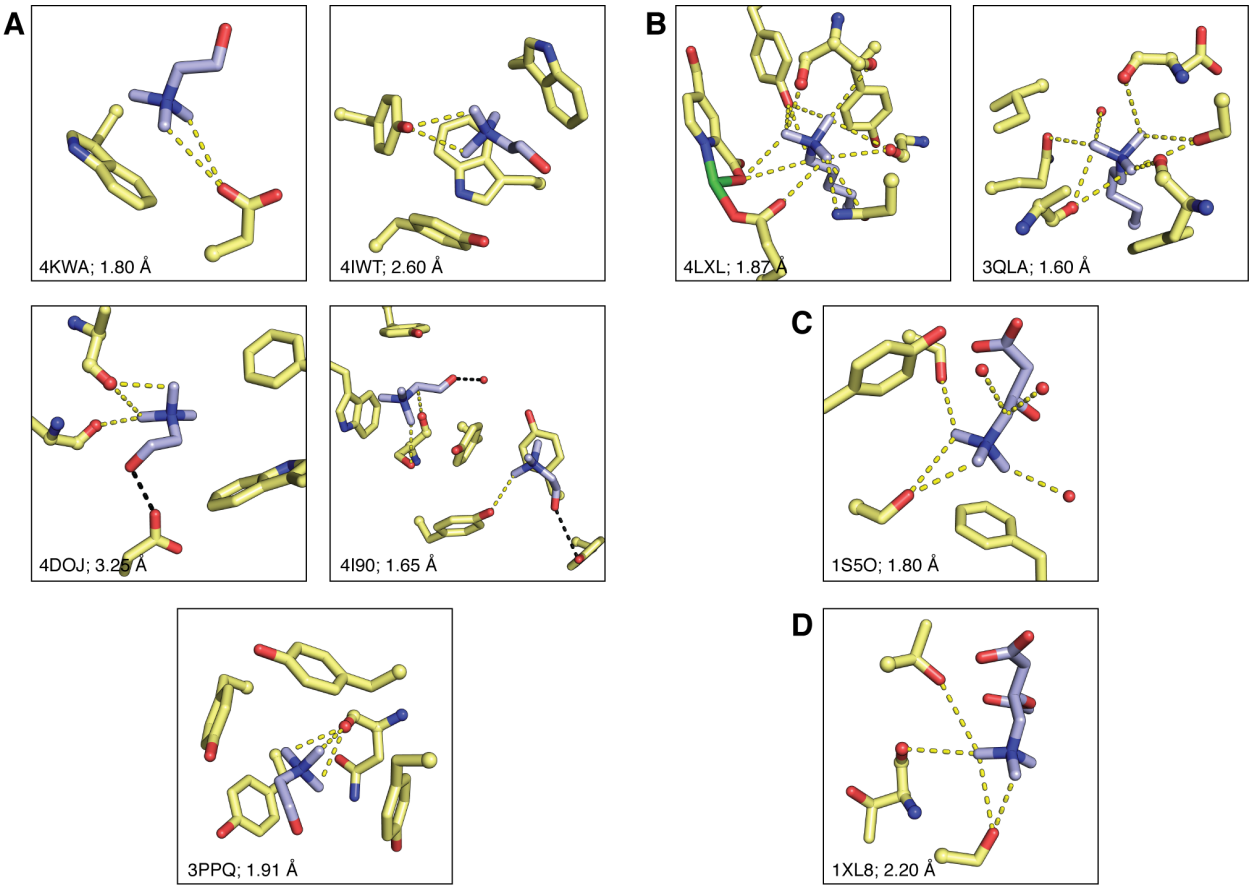


Figure 4.6: CH–O interactions are common in quaternary-ammonium-binding proteins. **(A)** Structures of proteins in complex with choline: PDB ID 4KWA, putative *Saccharomonospora viridis* transcriptional regulator; PDB ID 4IWT, choline binding domain of lytic amidase LytA; PDB ID 4DOJ, betaine transporter BetP; PDB ID 4I90 phosphatidylinositol-specific phospholipase C; PDB ID 3PPQ, ABC transporter solute binding domain OpuC. **(B)** Trimethyllysine-binding proteins: PDB ID 4LXL, JmjC domain-containing histone demethylation protein 3B bound to histone H3, K9-trimethyllysine and pyridine-2,4-dicarboxylic acid coordinated to a Ni^{2+} ion; PDB ID 3QLA, ADD domain of transcriptional regulator ATRX with bound histone H3, K9-trimethyllysine. **(C)** PDB ID 1S5O, carnitine bound to human carnitine acetyltransferase. **(D)** PDB ID 1XL8, octanoylcarnitine bound to murine carnitine octanoyltransferase. PDB IDs and reported resolution are listed within each panel. All contacts closer than 3.7 Å were taken to indicate CH–O interactions. The average C—O distance observed was 3.4 ± 0.2 Å. The median number of CH–O interactions per structure was three, with a large variation.

4.2.3: Comparison of CutC structure with its homology model reveals the difficulty in predicting active site features in GREs

As detailed in Chapter 3, a CutC homology model was previously created based on the structure of GDH (37% sequence identity) (9). However, we found by elucidating the crystal structure of CutC in complex with choline that the secondary structural elements that line the CutC and GDH active sites are more divergent than expected (**Figure 4.7**). This divergence is due in part to unusual features associated with two of the β-strands adjacent to the active site: a short helix interrupts β1 and a bulge interrupts β6. Because the small helix has one turn in CutC instead of two as predicted, the homology model for CutC is shifted out of register such that Tyr208 is pointing out of the active site in the model, whereas in the crystal structure, this residue points into the substrate-binding pocket and interacts with choline (**Figure 4.7.B, C**). In contrast, the position of Asp216 on one of the sheet-like regions of β1 is correctly predicted. Within β6, Thr502 and Tyr506 are placed in essentially the correct position, but the backbone conformation of the loop between these residues is different, with the result that Ser503 is interacting with choline in the model, whereas X-ray data place it out of the active site. Choline was docked into the homology model and its positioning largely matched that observed in the crystal structure, with the exception of the hydroxyl group location (up in the model toward Ser503 and Tyr506 instead of pointing down toward the amide nitrogen of Cys489 in the crystal structure) (**Figure 4.7.D**). Notably, the absence of Tyr208 from the model's active site and the inability to computationally predict CH–O interactions left the upper face of the trimethylammonium moiety of choline completely without protein contacts, whereas in the structure, this group is tightly held by multiple interactions with CutC. It thus appears that despite the high identity between CutC and GDH enzymes (37 %), homology modeling of CutC based on the crystal structure of GDH missed a number of key active site features, a problem that may turn out to be common for other GREs. A previous bioinformatic study (22) identified a large number of GDH- and PFL-like genes that are 30-45 % identical to known enzymes, but lack conservation of residues within the Cys loop, and thus would likely not possess a known function. For these reasons, more effective computational tools are needed for the structural modeling and active site prediction of GREs.

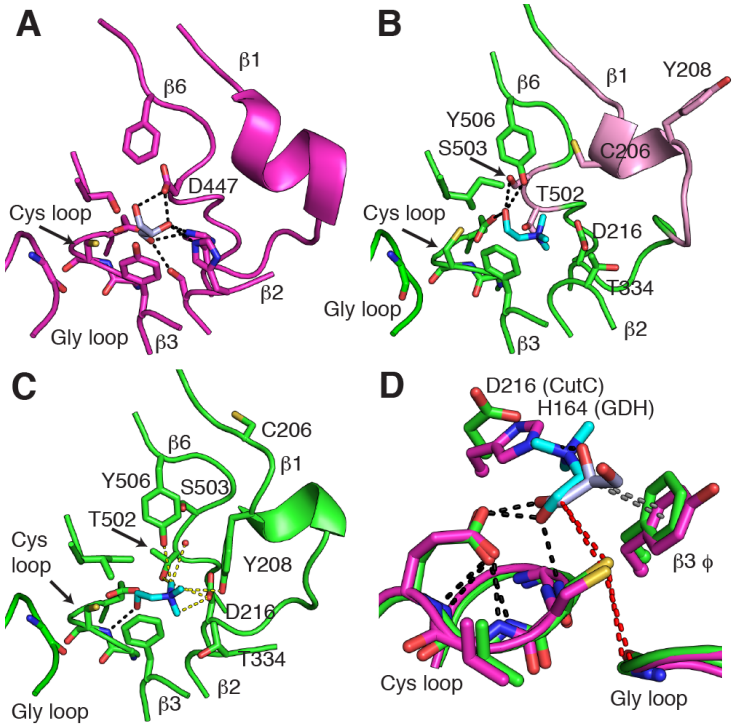


Figure 4.7: Comparison of CutC structure with GDH-based CutC homology model. The orientation of (a)-(c) is ~90º from that shown in Figure 4.4. **(A)** GDH structure with bound glycerol (gray). **(B)** Homology model of CutC created in Modeller using GDH as the starting model; choline (cyan) was docked using Schrödinger Suite 2012 (9). Residues that differ substantially from the observed positions in the crystal structure are highlighted (pink). **(C)** The crystal structure of CutC bound to choline (cyan) in the same orientation as in (b). CH–O interactions (yellow dashes) and hydrogen bonds (black dashes) are shown as in Figure 3. **(D)** An overlay of CutC (green) and GDH (magenta) with residues involved in binding C1 and C2 (sticks); the orientation is similar to that in Figure 4.4. The Cys loop GCVEP motif is conserved between the two enzymes. An aromatic residue contributed from β3 coordinates C2 (gray dashes). The proposed hydrogen transfer pathways (red dashes) between the active site Gly, Cys, and C1 are virtually identical in the two structures.

4.2.4: Disrupting enzyme-substrate CH–O interactions impairs binding and catalysis

Based on the crystal structure of CutC, Tyr208 and Tyr506 mutants were constructed to determine if the proposed CH–O interactions are important for binding choline or C–N bond cleavage (**Table 4.1**). The activation of these mutants by CutD occurred to a lesser extent than for wild-type CutC, perhaps due to disruptions in the conformational change required for glycyl radical installation or in the ability of the enzymes to stabilize and shield the glycyl radical from solvent. The fraction of the enzyme in radical form, however, appears stable over time (**Figure 4.8.A**). Kinetic analyses revealed retention in

affinity for choline in both CutC-Y208F and CutC-Y506F mutants, but a significant decrease (~57-fold) in the catalytic activity of CutC-Y208F. Given that Tyr208 establishes closer CH–O interactions (~3.3 Å) with the trimethylammonium group than Tyr506 (~3.5 Å), the greater impairment in activity for CutC-Y208F speaks to the importance of these contacts in properly positioning the backbone of choline for C–N cleavage. Interestingly, the double mutant CutC-Y208F/Y506F has pronounced defects in both binding (~12-fold increase in K_M) and activity (~83-fold decrease in k_{cat}).

Table 4.1: Activation and kinetic parameters of CutC mutants. The values reported with standard deviation are the average of three replicates. Activity for the mutants was initially detected by analysis of TMA production in an overnight end point assay. With the exception of CutC-D216N, mutants with detectable activity were kinetically characterized using a coupled assay as described in the methods. The un-normalized k_{cat} is given based on the total concentration of CutC. A normalized k_{cat} was estimated by dividing k_{cat} by the measured radical concentration.

protein	radicals per polypeptide (%)	detectable activity (TMA)	K_M (mM)	k_{cat} (s^{-1})	glycyl-radical-normalized k_{cat} (s^{-1})
wild type	20.8	yes	0.13 ± 0.01	157 ± 2	377 ± 5
E491Q	3.9	no	ND	ND	ND
E491A	10.5	no	ND	ND	ND
T502A	<1	yes	ND	0.036 ± 0.003	ND
D216N	<1	yes	ND	ND	ND
Y208F	5.5	yes	0.22 ± 0.03	0.73 ± 0.01	6.6 ± 0.1
Y506F	1.9	yes	0.12 ± 0.03	3.27 ± 0.11	86 ± 3
Y208F/ Y506F	1.7	yes	1.53 ± 0.12	0.15 ± 0.01	4.5 ± 0.1

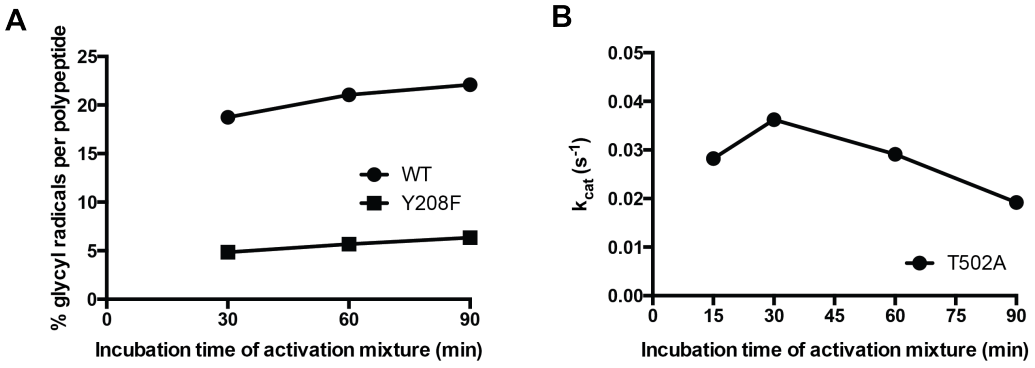


Figure 4.8: Glycyl radical stability of CutC-WT and mutants CutC-Y208F and CutC-T502A. **(A)** Stability of the glycyl radical signal for CutC-WT and CutC-Y208F was assessed by varying the incubation time of the activation mixture and measuring glycyl radical content by EPR. **(B)** The activity of CutC-T502A was used as a proxy for glycyl radical formation since the glycyl radical content of this protein was below the instrument's limit of detection. CutC-T502A was activated with CutD for the specified periods of time before recording the activity using a spectrophotometric coupled assay for acetaldehyde detection, as described in the methods. For each data set, the results of a representative experiment are shown.

To further investigate the contribution of CH–O contacts to choline binding, we crystallized CutC-Y208F after incubation with 10 mM choline. A mostly unperturbed active site is observed in the structure, with choline positioned similarly to wild-type CutC (**Figure 4.9.A**). The loss of the CH–O interactions between the phenolic hydroxyl and two of the methyl groups of choline does not substantially affect choline binding, in line with the relatively weak contribution of each CH–O contact to the overall binding enthalpy (30). A water molecule is found in one of the two molecules in the asymmetric unit in place of the phenolic hydroxyl, but close contacts with adjacent residues appear to preclude full occupancy, as the density for this water is weaker than those of other nearby waters. Presumably, this water substitutes for the phenolic hydroxyl in those active sites where it is present and could thus be important for catalysis in this mutant.

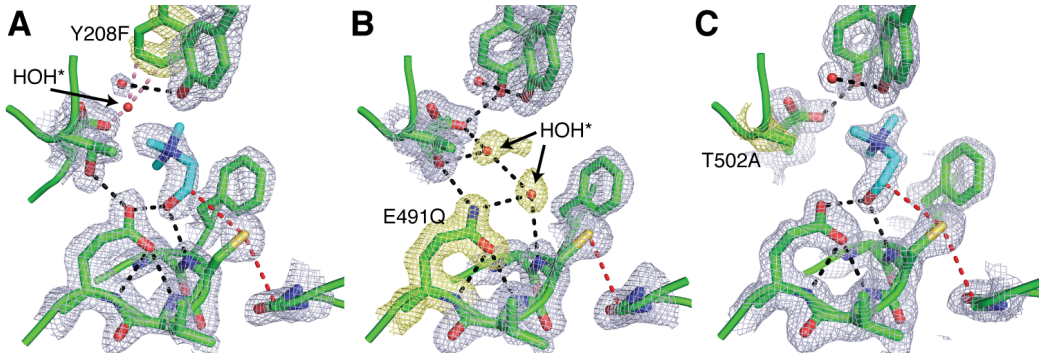


Figure 4.9: Structures of mutant CutC enzymes. **(A, B, C)** Composite omit density is contoured at 1.5σ around active site residues (light blue) with changes to each mutant structure highlighted (yellow). Hydrogen bonds (black dashes), steric clashes (pink dashes), and hydrogen-atom transfer pathways (red dashes) are indicated. Crystals were grown in the presence of 10 mM choline. **(A)** CutC-Y208F (1.90-Å resolution) closely resembles the wild-type enzyme, with the only difference being the generation of a new partially-occupied water binding site (HOH*). **(B)** The E491Q mutant (1.60-Å resolution) has no density for choline. Instead, two new water molecules are bound in the active site (HOH*). The orientation of the side chain amide is deduced based on the contacts to hydrogen bond donors within the Cys loop. **(C)** T502A at 1.85-Å resolution is virtually identical to the wild-type CutC structure.

4.2.5: Structure and mutagenesis support Glu491 as the putative catalytic base

The CutC mechanistic hypotheses in **Figure 4.1.C** and **4.1.D** require an acceptor for the choline hydroxyl proton, and Glu491 of the conserved GCVEP motif appears positioned to perform this function. Mutation of Glu491 would therefore be expected to have a large effect on substrate binding and catalysis. Indeed, we find that both CutC-E491Q and CutC-E491A are completely inactive in an overnight end

point assay (**Figure 4.10.A, Table 4.1**). To further investigate the source of the inactivity, we determined the crystal structure of CutC-E491Q. Mutant and wild-type enzymes are virtually identical in structure (root mean square deviation (RMSD) of 0.18 Å) including the arrangement of residues in the active site (RMSD of 0.09 Å for Cys loop residues). Despite inclusion of 10 mM choline in the protein solution prior to crystallization, no clear density is observed for choline within the active site (**Figure 4.9.B**). Instead, two new water molecules are bound: one replaces the hydroxyl group of choline and the other makes contacts to Asp216 and Thr502.

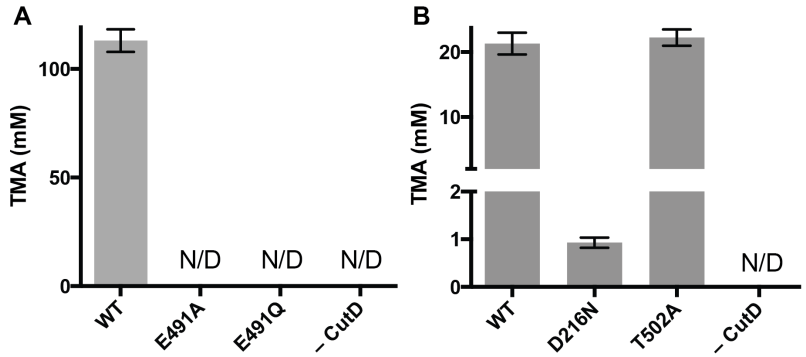


Figure 4.10: LC-MS end point assays quantifying TMA production by CutC wild-type and active site mutants upon overnight incubation. **(A)** Assays were conducted with 100 mM choline in phosphate buffer with CutC purified by Ni-affinity chromatography (as described in Online Methods). No TMA was detected in samples containing CutC-E491Q or CutC-E491A, nor in the sample lacking CutD (limit of detection ~1 μM). **(B)** Assays conducted with 20 mM choline in Tris buffer with CutC further purified by size exclusion chromatography (as described in methods). Each assay was performed in triplicate and the bar represents the mean ± standard deviation of three replicates. WT = wild type, N/D = not detected.

Analysis of predicted hydrogen orientations in wild-type CutC suggests that the choline C1 hydroxyl acts as a hydrogen bond donor to Glu491 and acceptor to the backbone nitrogen of the Gly488/Cys489 peptide bond (**Figure 4.11.A**). In the E419Q mutant structure, the side chain amide carbonyl maintains a hydrogen bond to a backbone amide nitrogen atom of the Cys loop. This orientation requires the side chain amide nitrogen, which is a hydrogen bond donor, to point toward the choline-binding site (**Figure 4.11.B**). Thus, the E491Q mutant lacks a hydrogen bond acceptor for the choline hydroxyl. We infer from the absence of density for choline in the active site of this mutant that CutC-E491Q has a greatly reduced affinity for choline. Thus, we cannot determine the severity of the defect in catalysis for E491Q, but given the conserved positioning of this residue and the complete absence of

activity even at high choline concentrations (100 mM), we propose that Glu491 is the catalytic base required for choline cleavage.

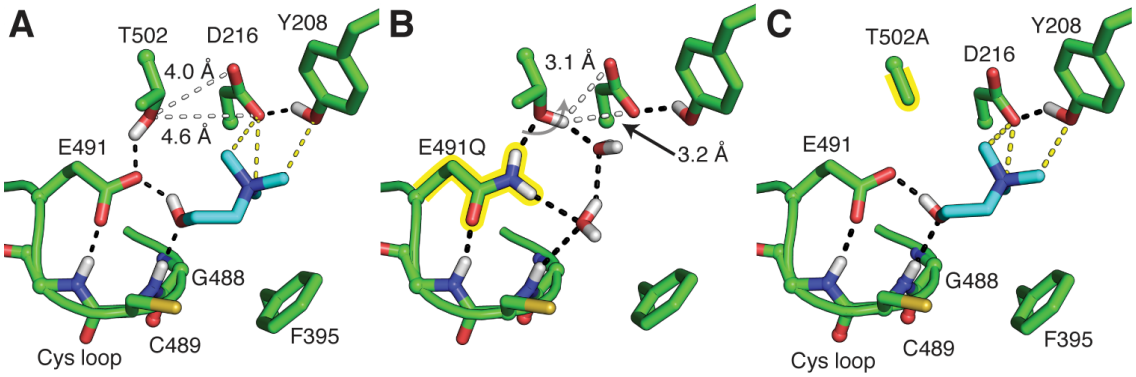


Figure 4.11: Putative hydrogen bonding networks in wild-type and mutant CutC enzymes. The orientation is shifted ~90º around the y-axis from Figure 4.4. Tyr506 and the Gly loop are not shown for clarity. Polar hydrogen atoms are shown within the networks based on inferred position given the available donor and acceptor groups, and considering that the pH of the crystallization was 8.0. Riding hydrogen atoms were added to the model during the final rounds of refinement. **(A)** In wild-type CutC, two networks of hydrogen bonds are highlighted (black dashes). Interactions between Glu491 and Thr502 and a backbone amide suggest that it is deprotonated and able to act as a general base to accept a proton from choline (cyan). Asp216 is also likely to be deprotonated, sharing a single proton in a hydrogen bond (O–O distance of 2.7 Å) with Tyr208. A proposed connection (white dashes) between Thr502-Asp216 would be possible if either of these residues were to change conformation. CH–O interactions to Tyr208 and Aps216 are shown as in Figure 4.4. **(B)** In CutC-E491Q, the side chain carbonyl of Q491 is likely to be pointing down toward the backbone amide, leaving the side chain nitrogen to point toward Thr502. **(C)** In CutC-T502A, no connection is available between Glu491 and Asp216.

4.2.6: A putative proton transfer network in the active site of CutC.

In order to complete the catalytic cycle, the protonated active site base Glu491 must be deprotonated. The CutC active site does not appear to be set up to allow transfer of protons to or from the exterior of the protein as there is no direct solvent pathway from the active site in any of our crystal structures. Instead we suggest that TMA deprotonates Glu491 for the next round of catalysis. Direct deprotonation of Glu491 by TMA would require a conformational change of CutC as the nitrogen is 4.6 Å away from the side chain of Glu491 prior to C–N bond cleavage and is sterically blocked by C1 of choline and Thr502 (**Figure 4.11.A**). While Thr502 establishes a close hydrogen bond with Glu491 (2.5 Å), the hydroxyl group of Thr502 is too far away (5.0 Å) from the trimethylammonium nitrogen and its pKa is likely too high to perform a concerted proton transfer during C–N cleavage. Instead, Asp216 is positioned close to the trimethylammonium moiety such that it could serve as the general acid in the

correct protonation state (**Figure 4.11.A**). However, Asp216 is also too far from Glu491 for direct proton transfer (5.8 Å), requiring a ~3 Å movement of the Asp216 side chain, which appears to be blocked by Thr502. Our structures suggest that Thr502 may bridge the gap between Glu491 and Asp216 by simultaneously accepting and donating a proton. Thr502 is already in position to hydrogen bond with Glu491 (2.5 Å), and a simple rotation of Thr502 around χ_1 by 10-15º should allow the formation of a new hydrogen bond between Thr502 and Asp216. In fact, in the structure of CutC-E491Q, which may serve as a mimic for the protonated state of Glu491, the hydroxyl group of Thr502 is positioned in hydrogen-bonding distance (3.6 Å) of Asp216 (**Figure 4.11.A, B**).

To further explore the roles of Asp216 and Thr502, which are both absolutely conserved in CutC enzymes (**Figure 4.5**), we generated CutC-D216N and CutC-T502A mutant proteins. The activation of both mutant proteins was below the limit of detection of our EPR instrument, but both enzymes were active in an overnight endpoint assay for TMA detection (**Figure 4.10.B**). Unfortunately, the low activity of CutC-D216N in the overnight assay precluded further kinetic analysis of this mutant. For CutC-T502A, a kinetic assay was performed and low activity was observed; however, this activity started to diminish after 30 min (**Table 4.1, Figure 4.8.B**). We were able to determine the structure of CutC-T502A, which shows choline binding at full occupancy with no discernable differences from wild-type (**Figure 4.11.C**), except for the fact that the hydrogen bonding network is interrupted (**Figure 4.11.A**).

4.2.7: Structure-based mechanistic proposals for CutC-catalyzed choline cleavage

The C–N lyase chemistry performed by CutC is unique among GREs, but resembles that of other radical enzymes that promote 1,2-elimination reactions: the C–O cleaving GRE GDH, the AdoCbl dependent enzymes EAL and diol dehydratase, and the AdoMet radical enzyme DesII. Based on previous studies of these enzymes, we have considered two mechanisms for CutC (**Figure 4.1.C** and **4.1.D**): a direct elimination of TMA from the substrate-based radical, similar to the one favored for GDH (14) and DesII (13), or a 1,2-migration, similar to the proposed mechanism for EAL (12) (**Figure 4.12**). The biochemical and structural data presented here appear to complement the elegant mechanistic studies performed with DesII using deuterated and fluorinated substrate analogs (32), and the structural and

computational studies for GDH (14, 17). As explained below, on the basis of these experiments, we favor a direct elimination mechanism for choline cleavage.

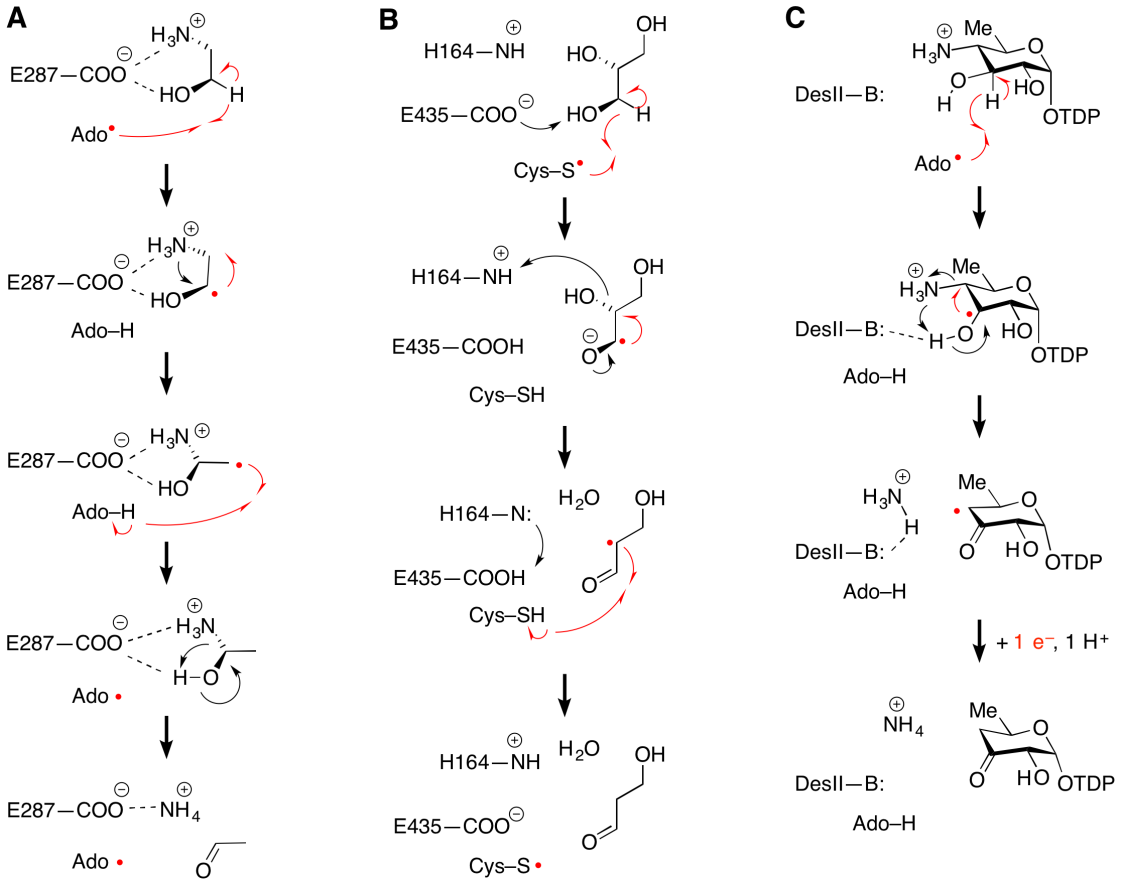


Figure 4.12: Proposed mechanisms for reactions catalyzed by enzymes described in the text. **(A)** Ethanolamine ammonia lyase (EAL) utilizes AdoCbl to generate a putative, transient Ado• intermediate which abstracts a hydrogen atom from the substrate, ethanolamine (12). A radical-assisted 1,2-migration of the amine group is facilitated by Glu287, which forms a bridging hydrogen bond between the alcohol and amine (33) (see Figure 4.13). The resulting carbinolamine radical regenerates a transient Ado• (recombines with Cbl to form AdoCbl) and subsequently decomposes to ammonium and acetaldehyde. **(B)** GDH, like CutC, utilizes a putative thiyl radical intermediate to abstract a hydrogen from its substrate, glycerol. Glu435 is proposed to serve as an active site general base that can deprotonate the resulting radical (17). His164 is the proposed active site general acid that facilitates dehydration by protonating the leaving hydroxide to form water (14). The resulting product radical can regenerate the transient thiyl radical. The active site acid and base are close enough to exchange a proton directly to reset the charge state of the active site (14). **(C)** DesII utilizes AdoMet and an $[4Fe\text{-}4S]^{+}$ cluster to generate a putative, transient Ado•, which removes a hydrogen from the substrate, TDP-D-quinovose (13). An unidentified active site general base is proposed to facilitate proton migration to the leaving amine, forming ammonium (32). We have chosen to illustrate two, distinct proton transfer steps, but a concerted elimination-protonation has also been suggested (32). The resulting product-based radical is reduced by an electron and protonated to form the deaminated product, TDP-3-keto-6-deoxy-D-glucose. Crucially, DesII does not regenerate AdoMet, but instead relies on an external source of reducing equivalents that has yet to be identified (34).

We propose that CutC-mediated catalysis is initiated by the formation of a transient thiyl radical (**Figure 4.13, I**). Even though the structures presented here are of unactivated CutC, the position of choline within the active site clearly supports the abstraction of a hydrogen atom from C1 (likely the pro-*S*) by the Cys489 thiyl radical as the probable first step in choline cleavage. This observation is supported by our experiments with C1-deuterium labeled choline analogs presented in Chapter 3. Hydrogen atom abstraction from choline by the putative thiyl radical could be facilitated by Glu491 hydrogen bonding to the C1 hydroxyl of choline. This hypothesis is based on a previous study of the C–D vibrational frequencies in primary and secondary alcohols, which determined that the strength of α C–H bonds of alcohols decreases upon increasing the strength of hydrogen bonding between the solvent and the hydroxyl group (35). It is important to note that the intermediacy of a thiyl radical in GRE catalysis has not yet been experimentally demonstrated. Nonetheless, our structures are similar to those of other GREs in further highlighting this key role of the thiyl radical.

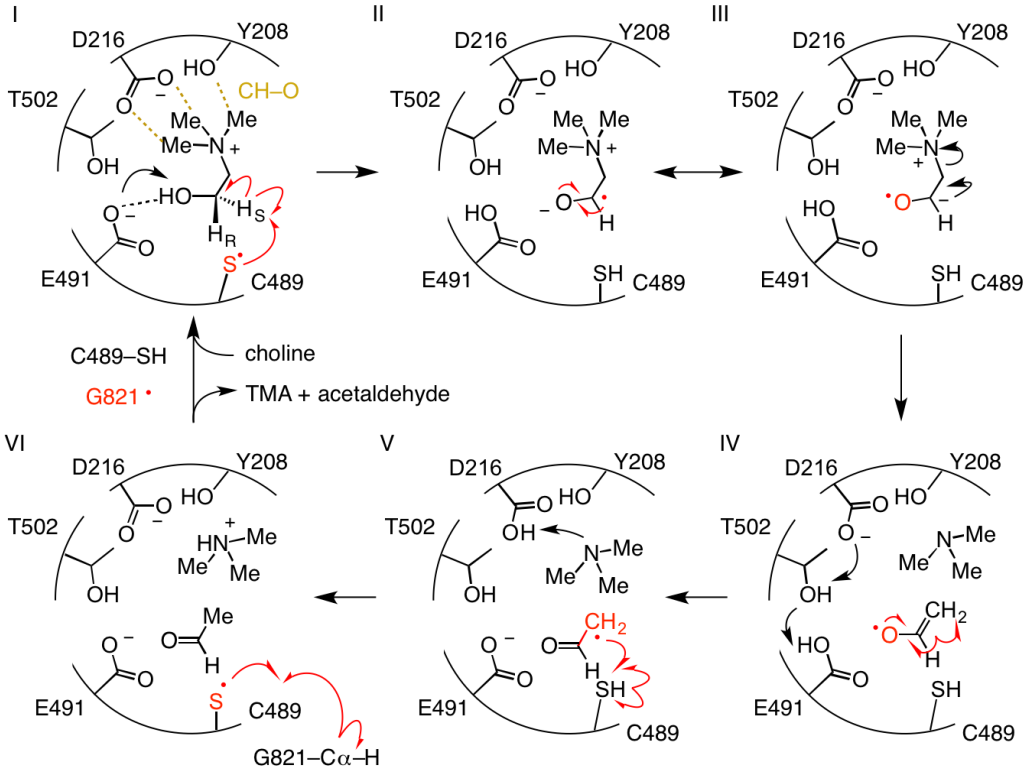


Figure 4.13: Mechanistic proposal for choline cleavage by CutC.

Our characterization of the E491Q and E491A mutants supports a role for Glu491 as a general base that could deprotonate a C1 choline radical to generate a α-hydroxyalkyl radical (**Figure 4.13, I to II**). Typically, the pKa of this radical species is 4-8 units lower than that of the corresponding alcohol (36) (pKa of the choline hydroxyl ~13.9 (37)), making this deprotonation step well within the capability of Glu491. This residue also occupies a position in the GRE fold (GCVEP) that is often associated with acid/base chemistry (25, 38) and is also conserved in GDH. A computational study of the GDH mechanism concluded that proton transfer between the C1 hydroxyl group of glycerol and an active site Glu residue triggers dehydratation via direct elimination (14). EPR and kinetic isotope effect studies of DesII also implicate a general base in the direct elimination of ammonia from TDP-4-amino-4,6-dideoxy-D-glucose, but the lack of a crystal structure for DesII precludes a specific assignment of this residue (13). A Glu is also present in the EAL active site, but its proposed role is different than in the other systems. By establishing close hydrogen bonds (2.8 Å) to both the ammonium and the hydroxyl moieties of ethanolamine, this residue may play a key role in stabilizing the transition state for the ammonium group migration (**Figure 4.14**). Thus, we propose that in CutC, Glu491 acts as a general base to form a transient ketyl radical, a species similar to those postulated in the 1,2-elimination of water by analogous radical dehydratases (39-41) and in the elimination of ammonia by DesII (13).

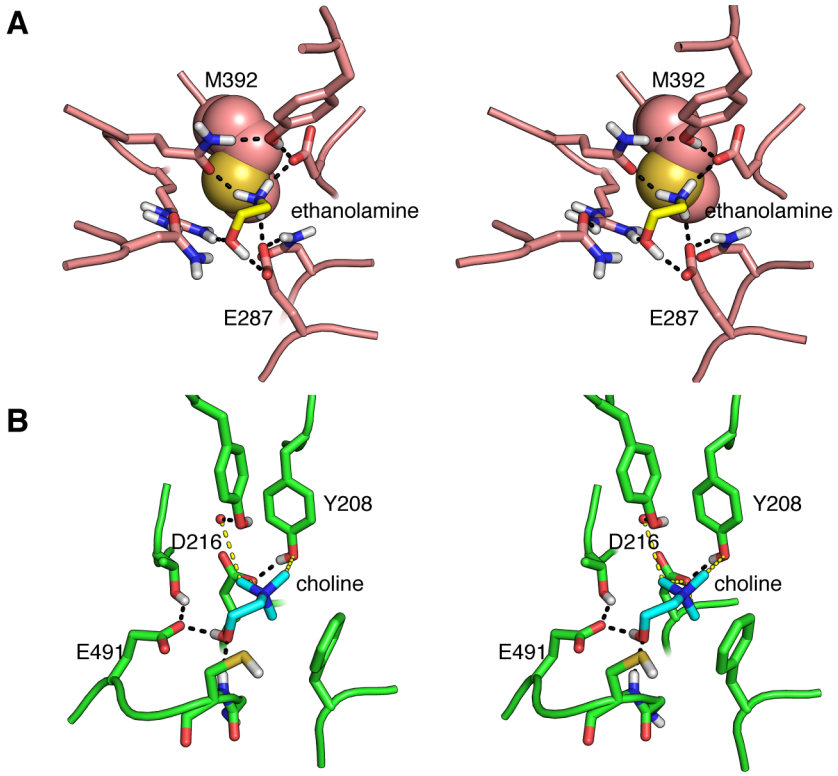


Figure 4.14: The CutC and EAL active sites are arranged to facilitate different chemistry. Protons were added to polar residues to make this figure using the H++ server (http://biophysics.cs.vt.edu/H++) (42); residues within 4 Å of the substrate are shown as sticks. Hydrogen bonds (black dashes) are shown between donor/acceptor pairs. **(A)** Stereoimage of the EAL active site with ethanolamine (yellow) bound (PDB ID 3ABO) (33). A 1,2 migration of the ammonium group is facilitated by Glu287, which forms hydrogen bonds with both the ammonium and the hydroxyl moieties of ethanolamine. A 1,2-elimination is disfavored by the side chain of Met392 (spheres), which is directly opposite the C–N bond, and sterically blocks the ammonium group from departing. Additionally, no suitable proton donor, such as Asp216 of CutC, is found adjacent to the substrate ammonium moiety. **(B)** The CutC active site, shown in a similar orientation to EAL relative to the substrate. CutC does not have a residue that is the equivalent of Glu287, which could assist in a migration reaction by forming hydrogen bonds with both the trimethylammonium and the hydroxyl group of choline. Additionally, there is no equivalent of Met392 to block elimination of the trimethylammonium, whereas there is a suitable proton donor (Asp216) for protonation of the trimethylamine leaving group.

Following deprotonation of the α-hydroxyalkyl radical, we propose that a rapid heterolytic C–N bond cleavage allows the direct elimination of TMA (**Figure 4.13, II, III**). This proposal draws support from radiolysis experiments in which choline-derived radicals undergo C–N fragmentation very rapidly under neutral conditions (pH 6.5–7.5) (43). EPR studies of the DesII-generated, substrate-based radical of TDP-D-quinovose have revealed the importance of establishing overlap between the partially filled p-orbital at the radical center and the σ* orbital of the adjacent functional group to be eliminated (44). Within the active site of CutC, choline is oriented such that the pro-*S* C1 hydrogen atom is antiperiplanar

to the trimethylammonium group, potentially allowing hyperconjugation between the p-orbital of a C1 choline radical and the C–N σ* orbital. The observed CH–O interactions between the trimethylammonium group and the polar residues lining the back of the active site (Asp216, Tyr208, Tyr506) would also be expected to weaken the C–N bond of choline, by aligning the trimethylammonium group for cleavage. Our structure suggests these CH–O interactions likely play a role in both choline binding selectivity and in catalysis, consistent with the ~83-fold reduced catalytic efficiency of the CutC double mutant Y208F/Y506F. Although perturbation of these CH–O contacts diminishes activity, this mutant is nonetheless active. In contrast, mutation of residues in EAL that interact with the ammonium group of ethanolamine, such as Glu287 (**Figure 4.14**), effectively abolish catalysis by EAL or lead to inactivation through formation of off-pathway radical intermediates (45). We ascribe this difference in the behavior of CutC and EAL mutants to the difference in proposed mechanism: facilitating a 1,2-migration reaction such as the one catalyzed by EAL requires that exquisite control be exerted over the migrating group by active site residues. No such requirement is necessary for a 1,2-elimination; indeed, this reaction proceeds very rapidly in solution (43). Taken together, this evidence highlights that even in its unactivated form, CutC binds choline in a configuration that could facilitate direct cleavage of the C–N bond from a ketyl radical.

Additional mechanistic steps to consider include deprotonation of the general base Glu491 to reset the CutC active site for the next round of catalysis (**Figure 4.13, IV**) and leaving group protonation (**Figure 4.13, V**). In GDH, the proposed general base for deprotonation of the C1 hydroxyl group is Glu435 (equivalent to Glu491), and the proposed general acid for protonation of the departing C2 hydroxyl group is His164 (analogous to Asp216) (**Figure 4.7.d**). Proton transfer between these two residues is thought to occur directly (14), with no equivalent of Thr502 blocking the requisite conformational changes. In CutC, Glu491 and Asp216 are too distant to interact directly and Thr502 appears to block the movement required for such interaction. Thus, we propose that the bridging/blocking residue Thr502 mediates proton transfer (**Figure 4.13, IV**), akin to what has been proposed previously for Ser and Thr residues in other systems (46, 47). We anticipate computational studies may provide

additional insights into this putative proton transfer network, which has been challenging to study experimentally.

A subsequent mechanistic step (**Figure 4.13, V**) is re-abstraction of the hydrogen atom from Cys489 by the product-based radical species to reform the thiyl radical and generate acetaldehyde. For AdoCbl enzymes, this re-abstraction step to reform the 5′-deoxyadenosyl radical (5′-dA•) is energetically challenging. It has been proposed for EAL that 1,2-migration facilitates re-abstraction of a hydrogen from 5′-dA, as this step was calculated to be exothermic by 1.2 kcal/mol (48). If base-catalyzed direct elimination were to occur in AdoCbl enzymes, the vinoxy radical formed would be stabilized by resonance (C–H bond enthalpy, ~95.5 kcal/mol (49)) and could not effectively regenerate 5′-dA• (5′-dA–H bond enthalpy, 99.9 kcal/mol (50)). In contrast to 5′-dA•, the glycyl and thiyl radicals in GREs are much more easily regenerated by product radicals (Cys S–H bond enthalpy, ~87 kcal/mol (51)), removing the necessity for the migration mechanism.

4.2.8: Conclusions

With these structural and biochemical studies, we are beginning to develop a molecular view of how TMA is generated by CutC. Nonetheless, the mechanism of choline cleavage by CutC is still an open question, that could be perhaps answered using choline mimics (e.g., vinylcholine) that are converted into stabilized radicals upon C–N cleavage, as observed in the case of B_{12}-dependent glycerol dehydratase inhibitor but-3-ene-1,2-diol (52). Given that a small molecule inhibitor of microbial TMA production was recently reported to diminish atherosclerotic plaque formation in mice (15), understanding TMA biogenesis is both timely and important. By dissecting the mechanistic details of CutC catalyzed C–N cleavage and elucidating the importance of active site residues for binding and catalysis, our work will facilitate the rational design of specific inhibitors for choline TMA-lyase. Such molecules could achieve target specificity either by mimicking the binding mode of choline or by intercepting the proposed radical intermediates.

4.3. Materials and methods

4.3.1. Materials and general methods

All chemicals and solvents were purchased from Sigma-Aldrich, except that Luria-Bertani Lennox (LB) medium was purchased from (EMD Millipore). Acetonitrile and water used for liquid chromatography-mass spectrometry (LC-MS) were B&J Brand high-purity solvents (Honeywell Burdick & Jackson). Oligonucleotide primers for mutagenesis (**Table 4.2**) were synthesized by Integrated DNA Technologies. Recombinant plasmid DNA was purified with a Qiaprep Kit from Qiagen. DNA sequencing was performed by Beckman Coulter Genomics. Multiple sequence alignments were performed with Clustal W (53), and the results analyzed with Geneious Pro 6.1.6 (54). Nickel-nitrilotriacetic acid agarose (Ni-NTA) resin was purchased from Qiagen. Protein gels were purchased from BioRad. Size exclusion chromatography experiments were performed on a BioLogic DuoFlow Chromatography System (Bio-Rad) equipped with a Superdex 200 Increase 10/300 GL column (GE Healthcare) for analytical size exclusion or with a HiLoad 16/600 Superdex 200 (GE Healthcare) for preparative-scale size exclusion chromatography. Protein concentrations were determined according to the method of Bradford using bovine serum albumin (BSA) as a standard (55), or using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific) for CutC wild-type and mutants (ε = 128,870 M^{-1} cm^{-1} for WT, T502A, E491Q, E491A, D216N, ε = 127,380 M^{-1} cm^{-1} for Y506F, Y208F and ε = 125,890 M^{-1} cm^{-1} for Y208F/Y506F). The extinction coefficient for CutC was estimated using the ExPASy ProtParam tool (http://www.expasy.org/tools/protparam.html). Assays were conducted in a glove box (MBraun) under a N_2 atmosphere with less than 5 ppm O_2. Sonication for disrupting *E. coli* cells overexpressing CutD was performed with a Branson Sonifier S-450D equipped with a ½ inch horn in a glove box (Coy Laboratory Products) under an atmosphere of 96% nitrogen and 4% hydrogen, at 4 ^{o}C. The spectrophotometric coupled assay for CutC activity was conducted at 22 ºC in 96-well plates using a PowerWave HT Microplate Spectrophotometer (BioTek). The absorbance of each well was monitored at 340 nm for 10 min, reads being 20 s apart. Path length correction was employed such that absorbance values were

reported for a path length of 1 cm. Data were analyzed using Gen5 software and kinetic parameters were calculated using GraphPad Prism6.

Table 4.2: Forward primer sequences used for site-directed mutagenesis. The second primer used was the reverse complement of the given sequence.

Name of target	Sequence (5′ to 3′)
CutC-Y208F	GTGTCCGACTGCTCCTTCCATGCGGTGAACGGC
CutC-D216N	GTGAACGGCGGCGGCAACTCCAACCCCGGTTAC
CutC-Y506F	TGGACCTCCACCGGCTTTACCCAGTGGCCCATC
CutC-T502A	GGTATAGCCGGTGGAGGCCCACTGGTACAGACG
CutC-E491A	GGGTTGCGTGGCGCCGCAGAAATCA
CutC-E491Q	ACGGCCTGATTTCTGCGGTTGCACGCAACCCATCAG

4.3.2: Construct design, expression, and purification of CutC variants for crystallization

Plasmids containing the wild-type *D. desulfuricans* G20 *cutC* gene or the same gene with an 18- or 52-residue, N-terminal truncation of the resulting protein sequence were constructed and assayed previously (9). The N-terminal truncations remove a putative microcompartment localizing sequence previously noted to induce aggregation in microcompartment-localized EAL (56). Both 18- and 52-residue truncation constructs incorporate an N-terminal hexahistidine tag and thrombin cleavage site. Full-length and truncated constructs have similar activity, but the truncated proteins are more soluble and less prone to aggregation (9). The 52-residue truncated construct was favored for crystallization of active site mutants as it removes additional residues that are not conserved and that appear to be unstructured in the crystal structure of the 18-residue truncated construct. Site-directed mutagenesis to introduce the E491A, E491Q, Y208F, D216N, T502A, Y506F, and Y208F/Y506F mutations was performed as previously described (9). For assays, CutC wild-type and mutants were purified as previously described with phosphate buffer pH 8, but they were not concentrated after dialysis because the process of concentration after Ni-affinity chromatography was found to contribute to aggregation. In order to determine whether purification by size exclusion chromatography improved activity, CutC wild-type, CutC-T502A and CutC-D216N were fractionated on a HiLoad 16/600 Superdex 200 column using 50 mM potassium phosphate buffer pH 8, 50 mM KCl, 10% v/v glycerol. The fractions corresponding to the dimer were combined and concentrated. CutD was also purified as previously described, with the

exception that the cells were disrupted using a sonifier inside a 4 °C oxygen-free chamber (9). All proteins were judged to be of high purity by denaturing polyacrylamide gel electrophoresis (**Figure 4.15**).

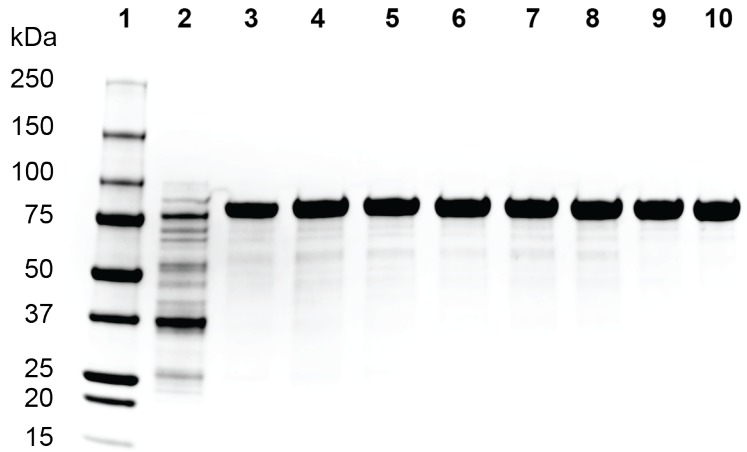


Figure 4.15: Sodium dodecylsulfate denaturing electrophoresis (4-15% Tris-HCl polyacrylamide gel) of the proteins used in this work. Precision Plus Protein All Blue Standards (BioRad) (lane 1), CutD (lane 2), CutC-WT (lane 3), CutC truncated mutants: CutC-E491A (lane 4), CutC-E491Q (lane 5), CutC-Y208F (lane 6), CutC-Y506F (lane 7), CutC-Y208F/Y506F (lane 8), CutC-T502A (lane 9), CutC-D216N (lane 10).

4.3.3: Crystallization and structure determination of wild-type CutC

Initial screening of non-post-translationally modified wild-type CutC protein with the intact N-terminal hexahistidine tag and 18-residue truncation was performed exposed to air with the aid of an Art Robbins Phenix micro-pipetting robot and a Formulatrix Rock Imager. Numerous initial conditions were found, with optimization yielding a well solution containing 17% (w/v) polyethylene glycol (PEG) 8000, 0.3 M lithium chloride, and 0.1 M Tris pH 8.0. Diffraction-quality crystals were obtained in hanging drop vapor diffusion trays at 21 ºC. Protein at 15 mg/mL in buffer containing 50 mM potassium phosphate pH 8.0, 50 mM KCl, and 10% (v/v) glycerol was mixed with well solution in a 1:1 ratio. Plate-like crystals formed after 2-3 days of equilibration and grew to maximum size over one week. Crystals were cryoprotected by soaking for 15-30 min in solution containing 20% (v/v) glycerol, 25% (w/v) PEG 8000, 0.5 M lithium chloride, and 0.1 M Tris pH 8.0. Choline was not added to the protein prior to crystallization but was included in the cryoprotection solution at 10 mM.

Crystals were initially characterized using an in-house CuKα X-ray source (Rigaku) and indexed in space group $P2_1$ with cell edges a = 79.6 Å, b = 234.9 Å, c = 105.0 Å, β = 109.6º. High-resolution

diffraction images were collected at the Advanced Photon Source beamline 24ID-C at a wavelength of 0.9795 Å on a Pilatus 6M detector (Dectris). Data were indexed, integrated, and scaled in HKL2000 (57) (statistics in **Table 4.3**). The structure of CutC was solved by molecular replacement in the Phenix implementation of Phaser (58). The structure of AdoCbl-independent glycerol dehydratase (PDB code 1R8W, 37% identity) (17) was used to construct a search model after trimming of non-identical side chains in Phenix Ensembler (59). A solution with four molecules per asymmetric unit was found with an initial R_{free} of 0.48 at 2.8-Å resolution. Several rounds of initial refinement in phenix.refine (59) with tight NCS restraints were sufficient to reduce R_{free} values below 0.4, and manual building of side chains further reduced R_{free} to ~0.3. NCS restraints were removed after initial refinement and addition of water molecules. Inclusion of explicit riding hydrogen atoms once the model was close to convergence improved both R_{free} and the model geometry. Positional and B factor refinement continued at the full resolution until the model was complete (statistics in **Table 4.3**). Choline was fit into difference density and verified with simulated annealing composite omit maps. Parameter files for choline were generated in Phenix eLBOW. Water molecules were placed automatically after ligands were refined and were verified manually. The final model contains all native residues at the C terminus in all four molecules. No density is observed for the N-terminal hexahistidine tag or residues 19-52 in the 18-residue truncated construct, but all other residues (53-846) are present and well ordered in the model. Despite different crystal contacts for each molecule in the asymmetric unit, the structures of each monomer were essentially indistinguishable, with root-mean-square deviations of 0.11-0.14 Å. Structural figures were made in PyMOL v1.4.1 (Schrodinger).

Table 4.3: Data collection and refinement statistics for wild-type CutC with choline.

Data collection and processing[a,b]	
Resolution (Å)	50-1.90 (1.93-1.90)
Completeness (%)	99.3 (99.1)
$\langle I/\sigma I \rangle$	12.5 (1.60)
R_{sym}	0.09 (0.99)
R_{pim}[c]	0.037 (0.430)
Unique reflections	282491

Redundancy	6.7 (5.9)
CC1/2	(0.68)
Model refinement	
Resolution (Å)	50-1.90 (1.92-1.90)
R_{work}, R_{free}	0.14 (0.18)
R_{free} set size	8478
Protein atoms[d]	25099
Choline atoms[d]	28
Other ligand atoms[d,e]	48
Water molecules	2486
Riding hydrogen atoms	24705
RMSD[f] Bond length (Å)	0.006
RMSD Bond angle (º)	0.999
Coordinate error (Å)	0.20
Rotamer outliers (%)	0.48
Ramachandran plot (%)	
Most favored	97.3
Additionally allowed	2.4
Disfavored	0.3
Average B factors (Å^2)[g]	
Monomer A	48.3
Monomer B	39.3
Monomer C	36.9
Monomer D	35.1
Choline	32.3
Other ligands[e]	55.7
Water	42.9

[a] Data were trimmed based on CC1/2 value rather than R_{sym} or <I/σI>. Inclusion of highest resolution data improved the quality of the electron density maps.
[b] Values in parenthesis are for the highest resolution shell.
[c] R_{pim} is reported due to the high redundancy of this datasets.
[d] Non hydrogen atoms
[e] Glycerol is bound at the protein surface.
[f] RMSD, root-mean-square deviation
[g] All atoms

4.3.4: Crystallization and structure determination of CutC mutants

Initial attempts to crystallize CutC mutants in the same condition as wild-type CutC were unsuccessful. Therefore, new crystals of CutC mutants (E491Q, T502A, or Y208F) with an N-terminal 52-residue truncation were identified in screening trays using the same setup as for wild type. Optimized crystals were grown using hanging drop vapor diffusion at 21 ºC (room temperature). CutC mutant protein at 8 mg/mL in buffer containing 50 mM potassium phosphate pH 8.0, 50 mM potassium chloride,

and 10 mM choline was mixed in a 1:1 ratio with well solution containing 1.0-1.2 M sodium malonate pH 7.0-8.0. Crystals were rod-like and grew within 7 days. Crystals were cryoprotected by brief transfer into a solution containing 3.4 M sodium malonate pH 7.0 and 10 mM choline and cryocooled by plunging in liquid nitrogen.

Diffraction images were indexed in space group $P4_22_12$ with cell edges approximately a = b = 230 Å, c = 79 Å. All data were collected and processed as for wild-type CutC (statistics in **Table 4.4**). Initial data processing in space group $P4_2$ suggested the presence of merohedral twinning in all datasets with twin fractions near 0.5. Despite this complication, the structure of each mutant could be solved unambiguously and with reasonable phasing statistics by molecular replacement in space group $P4_22_12$ in the Phenix implementation of Phaser (58) using the wild-type monomer as a search model. Two molecules are present per asymmetric unit, but the dimer interface is crystallographic, resulting in an overall tetrameric structure similar to the wild type. Attempts at twin refinement in space group $P4_2$ resulted in strong phase bias, little improvement in R_{free}, and a refined twin fraction of 0.5. Thus, the structures were refined without consideration of twinning in space group $P4_22_12$. Positional and B factor refinement was conducted with phenix.refine (59) (statistics in **Table 4.4**). Initially, refinement was performed only on the T502A structure, as the initial electron density maps were best for this dataset. Once the refinement converged, each mutant was refined separately (with R_{free} flags propagated to the other datasets). NCS and reference model restraints were used initially to reduce overfitting of the model but were removed after convergence with no change in the R_{free} values. Riding hydrogen atoms were included at the end of refinement, resulting in an improvement of R_{free} and geometry. Each final model contains native protein residues 53-846 in each chain; residues 1-52 were omitted in the construct. The N-terminal thrombin cleavage tag and linker are clearly visible in all chains; the hexahistidine tag is not ordered, but a single histidine residue is visible in one chain. Choline and active site waters were modeled into omit maps after refinement of the remainder of the protein. Three sodium ions and eight malonate ions are present in the final model, likely as a result of the high concentration of sodium malonate in the

crystallization solution. No sodium ions were found in the wild-type structure. All ligands and mutated residues were verified rigorously with composite omit maps.

Protein coordinates and structure factors have been submitted to the Protein Data Bank under accession codes 5FAU (wild type), 5FAV (E491Q mutant), 5FAW (T502A mutant), and 5FAY (Y208F mutant).

Table 4.4: Data collection and refinement statistics for mutant CutC proteins.

	CutC T502A choline	CutC E491Q	CutC Y208F choline
Data collection and processing[a,b]			
Resolution (Å)	50-1.85 (1.88-1.85)	50-1.60 (1.63-1.60)	50-1.90 (1.93-1.90)
Completeness (%)	99.9 (99.9)	100 (100)	99.7 (99.6)
<I/σI>	15.3 (3.0)	13.4 (2.6)	12.0 (2.0)
R_{sym}	0.154 (0.984)	0.128 (>1)	0.124 (>1)
R_{pim}[c]	0.044 (0.301)	0.039 (0.382)	0.051 (0.443)
Unique reflections	177353	273075	163497
Redundancy	13.0 (11.4)	13.1 (12.2)	7.2 (6.6)
CC1/2	0.73	0.69	0.63
Model refinement[b]			
Resolution (Å)	50-1.85 (1.87-1.85)	50-1.60 (1.62-1.60)	50-1.90 (1.92-1.90)
R_{work}, R_{free}	0.171, 0.206	0.151, 0.178	0.159, 0.193
R_{free} set size	5327	8200	4901
Protein atoms[d]	12709	12780	12753
Choline atoms[d]	14	0	14
Other ligand atoms[d,e]	66	66	66
Water molecules	2374	2367	2372
Riding hydrogen atoms	12521	12572	12565
RMSD[f] Bond length (Å)	0.005	0.008	0.006
RMSD Bond angle (º)	0.901	1.121	0.925
Coordinate error (Å)	0.17	0.14	0.15
Rotamer outliers (%)	0.52	0.51	0.44
Ramachandran plot (%)			
Most favored	97.3	97.2	97.2
Additionally allowed	2.5	2.5	2.5
Disfavored	0.2	0.3	0.3
Average B factors (Å^2)[g]			
Monomer A	23.4	21.2	23.7
Monomer B	25.2	24.1	26.9
Choline	31.3	--	43.6
Water	37.1	37.4	39.9
Other ligands[d]	50.8	46.2	57.1

[a] Data were trimmed based on CC1/2 value rather than R_{sym} or <I/σI>. Inclusion of highest resolution data improved the quality of the electron density maps.
[b] Values in parenthesis are for the highest resolution shell.
[c] R_{pim} is reported due to the high redundancy of these datasets.
[d] Non-hydrogen atoms
[e] Malonate and sodium ions from the cryoprotectant are bound at the protein surface.
[f] RMSD, root-mean-square deviation
[g] All atoms

4.3.5: Size exclusion chromatography to verify oligomeric state of CutC wild type and mutants

A ~30 µM solution of CutC (wild type or mutant) purified with potassium phosphate buffer pH 8 was analyzed by size exclusion with a 50 µL injection volume. The proteins were eluted over 122 min with 50 mM potassium phosphate buffer pH 8, 50 mM KCl, 10% v/v glycerol at 0.25 mL/min. A solution of molecular weight markers (Bio-Rad, #151-190) analyzed under the same conditions was used to construct a standard curve. CutC mutants are dimeric in solution (**Figure 4.16**), consistent with previous results (9).

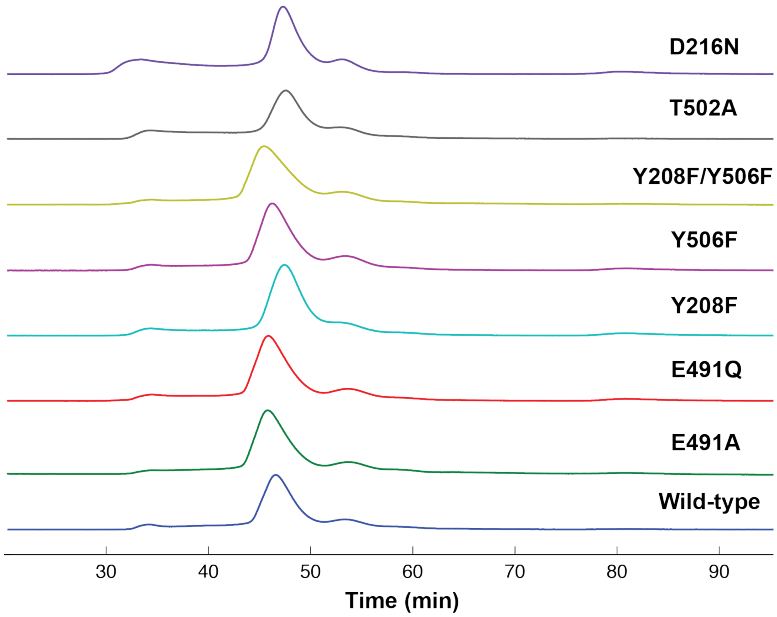


Figure 4.16: Determination of the oligomeric state of wild-type CutC and mutants by gel filtration chromatography. For this analysis, all proteins were purified by Ni-affinity chromatography. The major peak was found to correspond to the dimer.

4.3.6: Glycyl radical generation on CutC

Activation assays were set up in an MBraun glove box in 1.5 mL polypropylene Eppendorf tubes and contained 50 mM potassium phosphate pH 8, 50 mM potassium chloride, 150 µM sodium dithionite,

200 μM AdoMet, 10 μM CutC dimer, and 30 μM CutD in a total volume of 200 μL. A mixture of buffer, CutD and sodium dithionite was incubated for 20 min prior to addition of AdoMet and CutC, then glycyl radical formation was carried out for 1 h at room temperature. This incubation time was determined as optimal based on the stability of the glycyl radical signal for wild type CutC and CutC-Y208F (**Figure 4.8.A**). The conditions used for CutC activation were optimized as described below. Despite additional attempts at optimization, no glycyl radical signal was observed for CutC-T502A when the activation assay was incubated for 30 min or 60 min. X-band EPR spectroscopy and spin concentration measurements were performed as previously described (Chapter 3) (9), without taking into consideration the difference in receiver gain for standard and enzymatic assays as this difference is already accounted for by the spectrometer. The limit of detection for this technique was estimated at ~0.5 μM.

4.3.7: Optimization of glycyl radical formation on CutC

The activation assay for glycyl radical formation was optimized with wild-type CutC using a coupled, spectrophotometric assay for acetaldehyde, as described for steady state kinetics (section 4.3.8). The parameters surveyed and their respective outcomes are described in **Table 4.5**.

Table 4.5: Outcome of parameters surveyed for the optimization of glycyl radical formation on CutC.

Parameter surveyed for activation assay	**Outcome**
incubation time (30 to 90 min)	selected 60 min as incubation time
buffer (25 mM Tris pH 8 vs. 50 mM potassium phosphate pH 8)	selected 50 mM potassium phosphate pH 8 due to improved stability of the glycyl radical
ratio of CutC to CutD (from 1:0.4 to 1:4)	found that highest amounts of CutD afford best results, probably due to incomplete reconstitution of [4Fe-4S] clusters
using size exclusion chromatography-purified CutC	did not make a difference, but this step was found to decrease the stability of mutant CutC proteins
using size exclusion chromatography-purified CutD	did not make a difference, and avoiding this step might prevent potential oxygen damage of the [4Fe-4S] clusters
reductant (sodium dithionite (NaDT), *E. coli* flavodoxin (60) reduced with NaDT, acriflavin, methyl viologen, titanium (III) citrate)	moderate amounts of NaDT were found to afford best results
concentration of NaDT (2-20 equivalents with respect to CutC)	10-20 equivalents afforded best results
concentration of AdoMet (10-80 equivalents with respect to CutC)	20 equivalents afforded best results

For CutC-T502A, no glycyl radical signal was observed after 30 or 60 min incubation using the optimized activation conditions. In addition, no signal was observed after this mutant was purified by size exclusion chromatography, when Tris buffer was used instead of potassium phosphate, or when acriflavin (6 equivalents) was used instead of NaDT as reductant. For CutC-D216N, no EPR signal was observed when this mutant was assayed with Tris buffer after purification by size exclusion chromatography.

4.3.8: Kinetic analysis of choline cleavage

A spectrophotometric coupled assay monitoring oxidation of NADH by yeast alcohol dehydrogenase (YADH) was used to determine the kinetics of acetaldehyde production and thereby choline cleavage**.** Activation of CutC was carried out as described above. All assays contained 200 μM NADH, 50 mM potassium phosphate pH 8, 50 mM potassium chloride, diluted activation mixture (4000-fold for wild type, 33-fold for Y208F, 150-fold for Y506F, 10-fold for Y208F/Y506F, 2.5-fold for T502A), YADH (0.4-4 μM) and choline (0-10 mM, except for T502A, for which 100 mM choline was used) in a total volume of 200 μL. The assay was carried out in triplicate in a 96-well plate and the NADH absorbance at 340 nm was monitored for 7 min. The absorbance decreased linearly, indicating a constant concentration of activated enzyme during the course of our measurements.

4.3.9: LC-MS/MS assays for trimethylamine detection

For E491A and E491Q mutant proteins, glycyl radical formation on CutC was carried out on a 40 μL scale as described above, then 100 mM choline was added to the assays, followed by overnight (16 h) incubation in an Mbraun glove box at room temperature. For CutC-wild-type, CutC-T502A and CutC-D216N purified by size exclusion chromatography, activation assays were carried out in a similar manner, but contained 25 mM Tris buffer pH 8, 50 mM sodium chloride, 200 μM sodium dithionite, 200 μM AdoMet, 10 μM CutC dimer, and 40 μM CutD in a total volume of 50 μL. Choline was added to a final concentration of 20 mM, and the assays were incubated overnight (16 hr) in an Mbraun glove box at room temperature. For LC-MS/MS analysis, 4 μL of each assay were diluted into buffer R (20% 5 mM ammonium formate pH 4 and 80 % acetonitrile). This mixture was centrifuged at 13,000 rpm for 10 min to remove particulates, diluted 20 fold with buffer R, then 3 μL were analyzed by LC-MS/MS using a

previously-described method for underivatized TMA detection (precursor-product ion pair: m/z 60.1 → m/z 45, collision energy: 21 V, limit of detection after dilution: ~ 1 μM) (10).

4.4: References

1. Gaci N, Borrel G, Tottey W, O'Toole PW, & Brugere JF (2014) Archaea and the human gut: new beginning of an old story. *World J Gastroentero* 20(43):16062-16078.

2. Hippe H, Caspari D, Fiebig K, & Gottschalk G (1979) Utilization of trimethylamine and other *N*-methyl compounds for growth and methane formation by *Methanosarcina barkeri*. *Proc Natl Acad Sci USA* 76(1):494-498.

3. Krueger SK & Williams DE (2005) Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol Therapeut* 106(3):357-387.

4. Dumas ME*, et al.* (2006) Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci USA* 103(33):12511-12516.

5. Tang WH*, et al.* (2013) Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 368(17):1575-1584.

6. Wang Z*, et al.* (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57-63.

7. Christodoulou J (2012) Trimethylaminuria: An under-recognised and socially debilitating metabolic disorder. *J Paediatr Child Health* 48(3):E153-E155.

8. Craciun S & Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109(52):21307-21312.

9. Craciun S, Marks JA, & Balskus EP (2014) Characterization of choline trimethylamine-lyase expands the chemistry of glycyl radical enzymes. *ACS Chem Biol* 9(7):1408-1413.

10. Martinez-del Campo A*, et al.* (2015) Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *MBio* 6(2).

11. Selmer T, Pierik AJ, & Heider J (2005) New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem* 386(10):981-988.

12. Toraya T (2014) Cobalamin-dependent dehydratases and a deaminase: radical catalysis and reactivating chaperones. *Arch Biochem Biophys* 544:40-57.

13. Ruszczycky MW & Liu HW (2015) Mechanistic enzymology of the radical SAM enzyme DesII. *Isr J Chem* 55(3-4):315-324.

14. Feliks M & Ullmann GM (2012) Glycerol dehydratation by the B_{12}-independent enzyme may not involve the migration of a hydroxyl group: a computational study. *J Phys Chem B* 116(24):7076-7087.

15. Wang ZN*, et al.* (2015) Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell* 163(7):1585-1595.

16. Kalnins G*, et al.* (2015) Structure and function of CutC choline lyase from human microbiota bacterium *Klebsiella pneumoniae*. *J Biol Chem* 290(35):21732-21740.

17. O'Brien JR*, et al.* (2004) Insight into the mechanism of the B_{12}-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* 43(16):4635-4645.

18. Kerfeld CA, Heinhorst S, & Cannon GC (2010) Bacterial microcompartments. *Annu Rev Microbiol* 64:391-408.

19. Jorda J, Lopez D, Wheatley NM, & Yeates TO (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci* 22(2):179-195.

20. Becker A & Kabsch W (2002) X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA - how the enzyme uses the Cys-418 thiyl radical for pyruvate cleavage. *J Biol Chem* 277(42):40036-40042.

21. Vey JL*, et al.* (2008) Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA* 105(42):16137-16141.

22. Lehtio L & Goldman A (2004) The pyruvate formate lyase family: sequences, structures and activation. *Protein Eng Des Sel* 17(6):545-552.

23. Lehtio L, Grossmann JG, Kokona B, Fairman R, & Goldman A (2006) Crystal structure of a glycyl radical enzyme from *Archaeoglobus fulgidus*. *J Mol Biol* 357(1):221-235.

24. Martins BM*, et al.* (2011) Structural basis for a Kolbe-type decarboxylation catalyzed by a glycyl radical enzyme. *J Am Chem Soc* 133(37):14666-14674.

25. Funk MA, Judd ET, Marsh ENG, Elliott SJ, & Drennan CL (2014) Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proc Natl Acad Sci USA* 111(28):10161-10166.

26. Funk MA, Marsh ENG, & Drennan CL (2015) Substrate-bound structures of benzylsuccinate synthase reveal how toluene is activated in anaerobic hydrocarbon degradation. *J Biol Chem* 290(37):22398-22408.

27. Dougherty DA & Stauffer DA (1990) Acetylcholine binding by a synthetic receptor - implications for biological recognition. *Science* 250(4987):1558-1560.

28. Thompson JD, Higgins DG, & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22):4673-4680.

29. Horowitz S & Trievel RC (2012) Carbon-oxygen hydrogen bonding in biological structure and function. *J Biol Chem* 287(50):41576-41582.

30. Musah RA*, et al.* (1997) Variation in strength of an unconventional C-H to O hydrogen bond in an engineered protein cavity. *J Am Chem Soc* 119(38):9083-9084.

31. Horowitz S*, et al.* (2013) Conservation and functional importance of carbon-oxygen hydrogen bonding in Ado Met-dependent methyltransferases. *J Am Chem Soc* 135(41):15536-15548.

32. Lin GM, Ruszczycky M, Choi SH, & Liu HW (2015) Mechanistic investigation of the radical SAM enzyme DesII using fluorinated analogs. *J Am Chem Soc* 137(15):4964-4967.

33. Shibata N*, et al.* (2010) Crystal structures of ethanolamine ammonia-lyase complexed with coenzyme B_{12} analogs and substrates. *J Biol Chem* 285(34):26484-26493.

34. Ruszczycky MW, Choi SH, & Liu HW (2010) Stoichiometry of the redox neutral deamination and oxidative dehydrogenation reactions catalyzed by the radical SAM enzyme Desll. *J Am Chem Soc* 132(7):2359-2369.

35. Gawlita E*, et al.* (2000) H-bonding in alcohols is reflected in the Cα-H bond strength: variation of C-D vibrational frequency and fractionation factor. *J Am Chem Soc* 122(47):11660-11669.

36. von Sonntag C (2006) Free-radical-induced DNA damage and its repair: a chemical perspective (Springer).

37. Dawson RMC (1969) Data for biochemical research (Clarendon P., Oxford,).

38. Wei YF, *et al.* (2014) The class III ribonucleotide reductase from *Neisseria bacilliformis* can utilize thioredoxin as a reductant. *Proc Natl Acad Sci USA* 111(36):E3756-E3765.

39. Buckel W & Golding BT (1998) Radical species in the catalytic pathways of enzymes from anaerobes. *FEMS Microbiol Rev* 22(5):523-541.

40. Hans M, *et al.* (2002) Adenosine triphosphate-induced electron transfer in 2-hydroxyglutaryl-CoA dehydratase from *Acidaminococcus fermentans*. *Biochemistry* 41(18):5873-5882.

41. Lenz R & Giese B (1997) Studies on the mechanism of ribonucleotide reductases. *J Am Chem Soc* 119(12):2784-2794.

42. Anandakrishnan R, Aguilar B, & Onufriev AV (2012) H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* 40(W1):W537-W541.

43. Foster T & West PR (1974) Photolysis of aqueous solutions of hydrogen peroxide containing β-ammonio-alcohols. *Can J Chem* 52(21):3589-3598.

44. Ruszczycky MW, Choi SH, Mansoorabadi SO, & Liu HW (2011) Mechanistic studies of the radical S-Adenosyl-L-methionine enzyme DesII: EPR characterization of a radical intermediate generated during its catalyzed dehydrogenation of TDP-D-quinovose. *J Am Chem Soc* 133(19):7292-7295.

45. Mori K, *et al.* (2014) Catalytic roles of substrate-binding residues in coenzyme B_{12}-dependent ethanolamine ammonia-lyase. *Biochemistry* 53(16):2661-2671.

46. Erez Y, *et al.* (2011) Structure and excited-state proton transfer in the GFP S205A mutant. *J Phys Chem B* 115(41):11776-11785.

47. Patton GC, *et al.* (2011) Cofactor mobility determines reaction outcome in the IMPDH and GMPR (β-α)$_8$ barrel enzymes. *Nat Chem Biol* 7(12):950-958.

48. Wetmore SD, Smith DM, Bennett JT, & Radom L (2002) Understanding the mechanism of action of B_{12}-dependent ethanolamine ammonia-lyase: synergistic interactions at play. *J Am Chem Soc* 124(47):14054-14065.

49. da Silva G, Kim CH, & Bozzelli JW (2006) Thermodynamic properties (enthalpy, bond energy, entropy, and heat capacity) and internal rotor potentials of vinyl alcohol, methyl vinyl ether, and their corresponding radicals. *J Phys Chem A* 110(25):7925-7934.

50. Wetmore SD, Smith DM, Golding BT, & Radom L (2001) Interconversion of (*S*)-glutamate and (2*S*,3*S*)-3-methylaspartate: a distinctive B_{12}-dependent carbon-skeleton rearrangement. *J Am Chem Soc* 123(33):7963-7972.

51. Rauk A, Yu D, & Armstrong DA (1998) Oxidative damage to and by cysteine in proteins: An ab initio study of the radical structures, C-H, S-H, and C-C bond dissociation energies, and transition structures for H abstraction by thiyl radicals. *J Am Chem Soc* 120(34):8848-8855.

52. Pierik AJ, Graf T, Pemberton L, Golding BT, & Retey J (2008) But-3-ene-1,2-diol: A mechanism-based active site inhibitor for coenzyme B_{12}-dependent glycerol dehydratase. *Chem Bio Chem* 9(14):2268-2275.

53. Larkin MA*, et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.

54. Kearse M*, et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.

55. Stoscheck CM (1990) Quantitation of protein. *Method Enzymol* 182:50-68.

56. Akita K*, et al.* (2010) Purification and some properties of wild-type and N-terminal-truncated ethanolamine ammonia-lyase of *Escherichia coli*. *Journal of Biochemistry* 147(1):83-93.

57. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.

58. Mccoy AJ*, et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.

59. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66:213-221.

60. Hall DA, Jordan-Starck TC, Loo RO, Ludwig ML, & Matthews RG (2000) Interaction of flavodoxin with cobalamin-dependent methionine synthase. *Biochemistry* 39(35):10711-10719.

Chapter 5: Towards preventing the production of trimethylamine by the human gut microbiota

5.1: Introduction

It has long been recognized that gut bacteria play important roles in human health and physiology. Collectively possessing over 150-fold more genes than the human genome, our gut microbes have metabolic and biosynthetic capabilities that greatly surpass those of human cells (1). As described in Chapter 1 of this thesis, such functions can be beneficial to the host, since gut bacteria synthesize essential vitamin and nutrients, influence the development of immune cells, and digest inaccessible food components (2). The gut microbiota has also been implicated in a number of diseases, including inflammatory bowel disease (IBD), obesity, nonalcoholic fatty liver disease (NAFLD), and colon cancer (3). Despite great advances in our understanding of correlations between changes in the composition of gut microbial communities and host physiological states, the molecular mechanisms underlying host-gut microbiota interactions are still largely unknown. One approach to elucidate the impact of microbial metabolic activities on community structure, dynamics, and functions is the use of small-molecule inhibitors that target gut microbial enzymes. Currently, this approach was employed only with the purpose of blocking the activity of sugar scavenging gut bacterial β-glucuonidase enzymes, which deconjugate xenobiotics by removing glucuronide groups for use as carbon source (4). In the case of the cancer drug SN-38, this mode of deconjugation within the gut lumen results in severe diarrhea (4). Relying on an understanding of the structure and activity of bacterial β-glucuonidases, Wallace et al. uncovered inhibitors that selectively target microbial enzymes β-glucuonidases and were not toxic to either bacteria or human cells (4). Thus, this study opened the possibility of 'drugging the microbiota', *i.e.*, modulating the activity of specific microbial enzyme targets for scientific and clinical purpose (5).

Recently, a link has emerged between a metabolite of gut microbial origin, trimethylamine *N*-oxide (TMAO), and cardiometabolic disorders, such as NAFLD and atherosclerosis. As described in Chapter 1 of this thesis, the metabolite TMAO is primarily produced via a human-microbial co-metabolic pathway that begins with the conversion of the diet-derived nutrient choline to trimethylamine (TMA) by

gut microbes under anaerobic conditions (**Figure 5.1**). Phosphatidylcholine (PC) is believed to represent the main source of choline in the gut, as it is highly abundant in animal products and bile, and it can be converted into choline through the action of both human and microbial phospholipases (6). TMA is considered primarily to be a waste product of this metabolic pathway; bacteria utilize the resulting acetaldehyde as carbon source and only methanogenic Achaea present in low abundance in the human gut can further degrade TMA (7). As a result, most of the TMA produced in the gut is transported via the portal circulation to the liver, where flavin-dependent monooxygenase enzymes (FMOs), particularly FMO3, oxidize it to TMAO. Due to its high water solubility, TMAO is predominantly excreted by the kidneys into urine (8). Although TMA has been considered a normal constituent of human urine since the beginning of the past century, the exclusively microbial origin of this metabolite was not established until 1992 (9). As discussed below, due to recent reports linking high TMAO levels in urine and plasma to several diseases, the pathways leading to TMA and TMAO production have only lately been considered sources of new therapeutic targets (10).

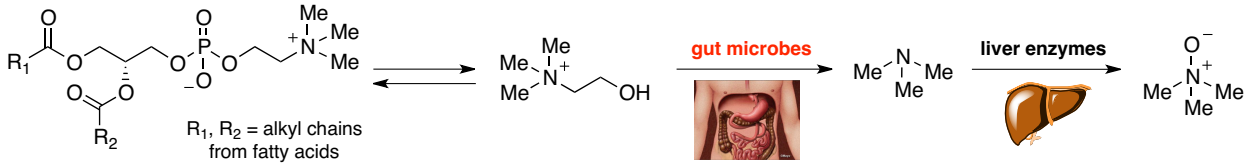


Figure 5.1: Overview of the human-microbial co-metabolic pathway for the conversion of choline into trimethylamine *N*-oxide (TMAO).

Multiple clinical studies have found that elevated plasma TMAO levels represent a prognostic marker for cardiovascular disease, independent of traditional risk factors, renal function, and medication use (11-13). Feeding TMAO or its main precursors (choline, phosphatidycholine and L-carnitine) to female atherosclerosis-prone ApoE$^{-/-}$ mice enhanced plaque formation (12, 13), and TMAO was found to alter cholesterol transport, induce inflammation in endothelial cells, and directly contribute to thrombosis by enhancing platelet aggregation (12-15). Since a molecular target for TMAO has not yet been discovered and the contribution of TMAO to atherosclerosis was contested by a different study (16),

chemical tools that inhibit the formation of this metabolite would be tremendously useful to dissect its contribution to disease.

Since TMAO is formed via a human-microbial co-metabolic pathway, recent studies investigated the effects of preventing its formation by flavin monoxygenase 3. Interestingly, FMO3 silencing using an antisense oligonucleotide in liver insulin receptor knockout (LIRKO) mice led to a decrease in circulating TMAO and a protective effect towards atherosclerosis development (17). However, it is difficult to decipher the cause of this outcome, since FMO3 knockdown also had a pronounced effect on lipid, glucose and insulin levels (17). These studies thus indicate that FMO3 has an important role in maintaining homeostasis, and that its inhibition via chemical tools is not a feasible approach to studying the role of TMAO on the progression of atherosclerosis. Instead, targeting TMA formation by gut microbes arises as a potentially more effective strategy.

Gut bacterial TMA production is also linked to liver diseases. It is well known that choline deficiency leads to fat accumulation in the liver (18); recently, a metabolomics study described a gut microbiota-mediated mechanism underlying the development of fatty liver in a mouse model susceptible to fatty liver disease and insulin resistance (19). Dumas et al. concluded that a gut microbiota abundant in choline-utilizing bacteria could reduce choline bioavailability and cause non-alcoholic fatty liver disease (NAFLD) (19). This conclusion was also corroborated by another study revealing that NAFLD is transmissible to germ-free mice via microbiota transplant (20). Even though NAFLD is considered the most common cause of liver disease worldwide and affects 17% of the world's population, therapeutic intervention strategies for this disease are limited to lifestyle changes, such as reducing caloric intake and increasing physical activity (21). Due to the contribution of gut microbial metabolism of choline to NAFLD, inhibitors of this microbial pathway would represent viable therapeutic options that could be part of a complex approach to treatment.

The inherited metabolic disorder trimethylaminuria (fish malodor syndrome) arises from mutations in the gene encoding FMO3 that result in an inability of individuals to completely oxidize TMA to TMAO, and an associated socially and psychologically debilitating malodor (22). Since many

people suffering from trimethylaminuria remain undiagnosed for long periods of time, the frequency of this disorder is not precisely known, but genetic studies have estimated the incidence of heterozygous carriers to range from 0.5% to 11% depending on the ethnicity examined (22). Several holistic approaches are known to manage the intense odor associated with this disease (as detailed in Chapter 1), but no universally efficacious treatment options for trimethylaminuria have been developed (22). Therapeutic intervention using a small molecule inhibitor selective for TMA production would thus be a highly beneficial strategy. Such molecules would be less likely to perturb the beneficial activities of the gut microbiota, or interact with mammalian enzymes, as demonstrated in the case of gut bacterial β-glucuronidase inhibition (4).

Prior to the discovery of the enzymes responsible for TMA formation by anaerobic microbes, several attempts were made to inhibit this activity in whole cell cultures (*Streptococcus sanguis* (23)) or cell lysates (*Desulfovibrio desulfuricans* (24, 25), *Clostridium* (26), *Proteus mirabilis* (27)) of anaerobic bacteria that converted choline into TMA and acetaldehyde. These studies determined that TMA production from choline was inhibited by 2,4-dintrophenol (mechanism still unknown (23-25, 27)), reagents that can react with cysteine residues (23-25, 27), iron-chelating reagents (24, 25, 27), a known choline uptake inhibitor (hemicholinium-3 (23)), and choline mimics (dimethylaminoethanol (23, 26) and betaine aldehyde (23)). Since the main purpose of these experiments was to gain information about the enzymes involved in TMA generation, the molecules listed above were not tested further to determine if they would be suitable for use with other strains of bacteria or in a host.

As described in Chapter 1 of this thesis, the genetic and biochemical bases of TMA production by gut microbes were unknown until recently, and anaerobic choline metabolism had not been extensively studied. Since the human colon is a predominantly anaerobic environment (28), the contribution of a recently discovered oxygen-dependent pathway (*yeaW/X*) for TMA formation from choline by gut microbes is likely very limited (29). As described in Chapter 2 of this thesis, we used a genome mining approach to uncover the first bacterial gene cluster (*cut* cluster) involved in anaerobic bacterial choline metabolism (**Figure 5.2**) (30). We also established that the *cut* pathway could be a major mechanism for

the direct conversion of choline to TMA by gut bacteria (31). The key TMA-generating enzyme, choline TMA-lyase (CutC), is a new member of the glycyl radical enzyme family, and represents a specific molecular marker for choline metabolism, as detailed in Chapters 3 and 4. In collaboration with a postdoctoral fellow in the Balskus group, Dr. Ana Martínez del Campo, members of Prof. Federico Rey's research group at the University of Wisconsin-Madison have also shown that organisms containing the *cut* genes, but not *cutC* mutants, produce TMA in gnotobiotic mouse models, including ApoE$^{-/-}$ mice (manuscript in preparation). This finding confirms that the *cut* pathway is functional in a host setting, and represents the first step towards investigating the impact of microbial choline utilization on host biology using tool compounds.

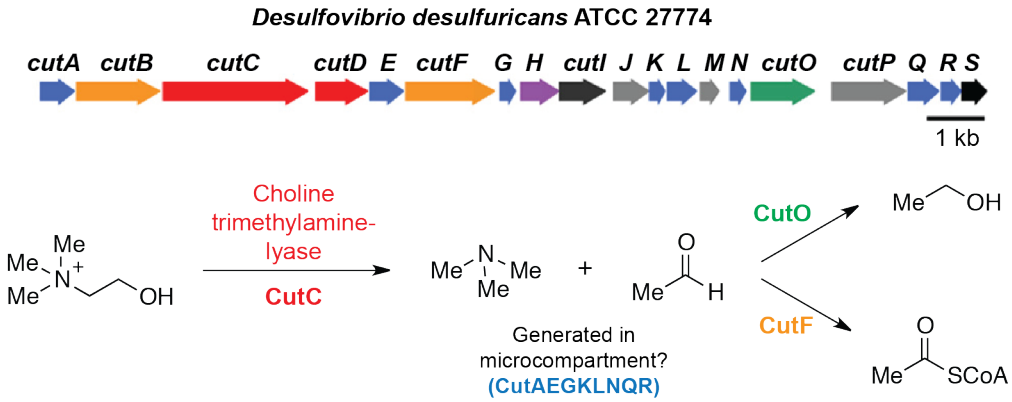


Figure 5.2: The choline utilization (*cut*) gene cluster and model for choline metabolism in *Desulfovbrio desulfuricans* ATCC 27774.

In this Chapter, I describe several strategies to discover small molecule inhibitors of gut microbial choline cleavage. An initial round of structure-guided inhibitor design afforded betaine aldehyde as a promising early lead. In vitro assays and X-ray crystallography indicated that this molecule specifically interacts with CutC, and could function as a reversible covalent inhibitor. Betaine aldehyde also inhibited the conversion of choline to TMA by a panel of various strains bacteria harboring the *cut* gene cluster and by a human fecal sample. I also describe the development of a high throughput screening approach to identify new inhibitors of choline metabolism by gut microbes, which could provide an orthogonal strategy to rational design approaches. Towards this end, a graduate student in the Balskus group, Marina Orman, and I optimized a media formulation containing choline as sole carbon source, ensuring that

bacterial survival would be dependent on the *cut* pathway. Under those conditions, betaine aldehyde inhibited choline-dependent but not glycerol-dependent growth of the *cut* gene cluster harboring strain *E. coli* MS 69-1. Overall, these preliminary results show that small molecules can interfere with anaerobic choline metabolism by gut microbes and set the stage for more extensive efforts to discover potent inhibitors this pathway.

5.2: Results and discussion

5.2.1: Rational design and testing of putative CutC inhibitors

Our first approach towards discovering small molecule inhibitors of anaerobic microbial choline metabolism involved the rational design of compounds that are structural mimics of choline or can interfere with the mechanism of CutC. To this end, we have leveraged our knowledge of the activity, mechanism, and structure of CutC (**Figure 5.3**). Our in vitro biochemical studies described in Chapter 3 revealed that the enzyme has high specificity for choline. We also confirmed the importance of two essential active site residues, the glycine that becomes a glycyl radical upon the action of the activating enzyme CutD, and a cysteine that would be converted into a thiyl radical upon substrate binding and could generate the choline-based radical at C1. The biochemical and structural studies described in Chapter 4 revealed that carbon-oxygen hydrogen (CH–O) interactions between the methyl groups of the trimethylammonium moiety of choline and O atoms of polar active site residues (Asp216, Tyr208, Tyr506) are important for both binding and catalysis. In addition, our insights enabled us to propose a mechanism for choline cleavage involving direct elimination of TMA from a substrate-based radical, upon deprotonation of the choline hydroxyl by Glu491 (**Figure 5.3.C**). Elucidating the molecular basis for choline binding and catalysis is extremely valuable for uncovering modulators of TMA production, providing us with the opportunity to perform rational design and structure-guided optimization of inhibitors.

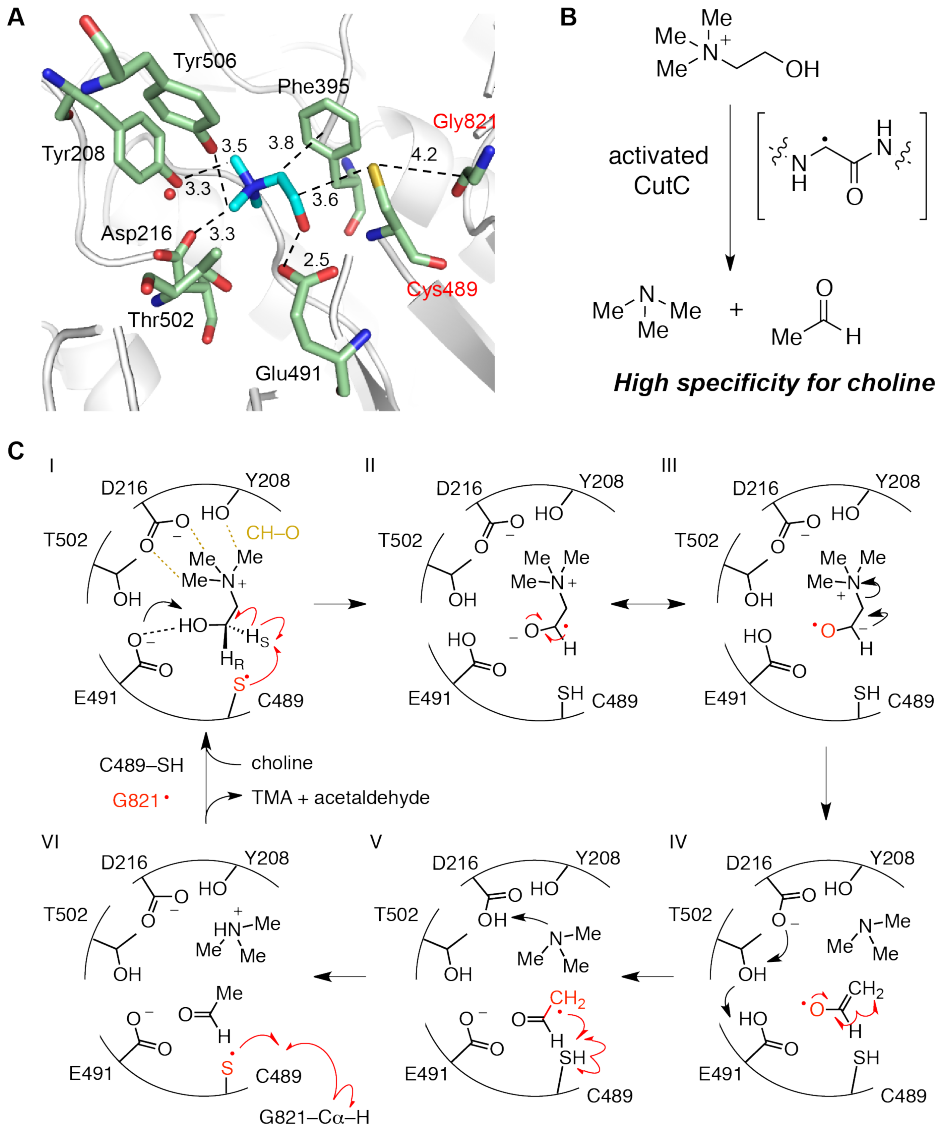


Figure 5.3: Structure and proposed function of choline TMA-lyase (CutC). **(A)** Choline-bound structure of CutC. **(B)** CutC utilizes a protein-based glycyl radical to perform its function. **(C)** Proposed mechanism for choline cleavage by CutC; for details, see Chapter 4.

For our inhibitor design we thus chose to focus on a variety of compounds similar in molecular weight to choline (**Figure 5.4.A**). We prioritized compounds that could potentially bind to other regions of the active site (compound **7**), establish similar interactions as choline with active site residues (compounds **3, 4, 5, 6, 8, 12**), engage the active site Cys residue (compounds **1, 2, 9, 10, 11, 13**), or stabilize the radical formed upon initial hydrogen atom abstraction from C1 (compounds **2** and **10**). We tested this small library of molecules (either purchased of synthesized as described in 'materials and methods') using in vivo assay for TMA-d_9 formation by *cut* cluster-containing bacteria *Escherichia coli*

MS 200-1 grown for 3.5 h in rich media (Brain Heart Infusion broth, BHI) supplemented with 1 mM (trimethyl-d_9)-choline (d_9-choline) (**Figure 5.3.B**). The inhibitors were added when the cultures were inoculated, and during this time interval, all cultures grew to saturation, indicating that none of the compounds tested possessed antibiotic properties. For reasons still under investigation, the extent of d_9-TMA production by *E. coli* MS 200-1 from the d_9-choline provided is ~30% under these conditions (3.5 h incubation), a value that matches the maximum possible conversion of choline into TMA (over 12 h) achieved by this strain of bacteria. Since the levels of d_9-choline and d_9-TMA at the end of our experiment add to approximately the concentration of d_9-choline initially added (1 mM), it does not seem that this strain of bacteria utilizes choline for other purposes, except in the case of compound **13**, in which choline might be diverted to other pathways.

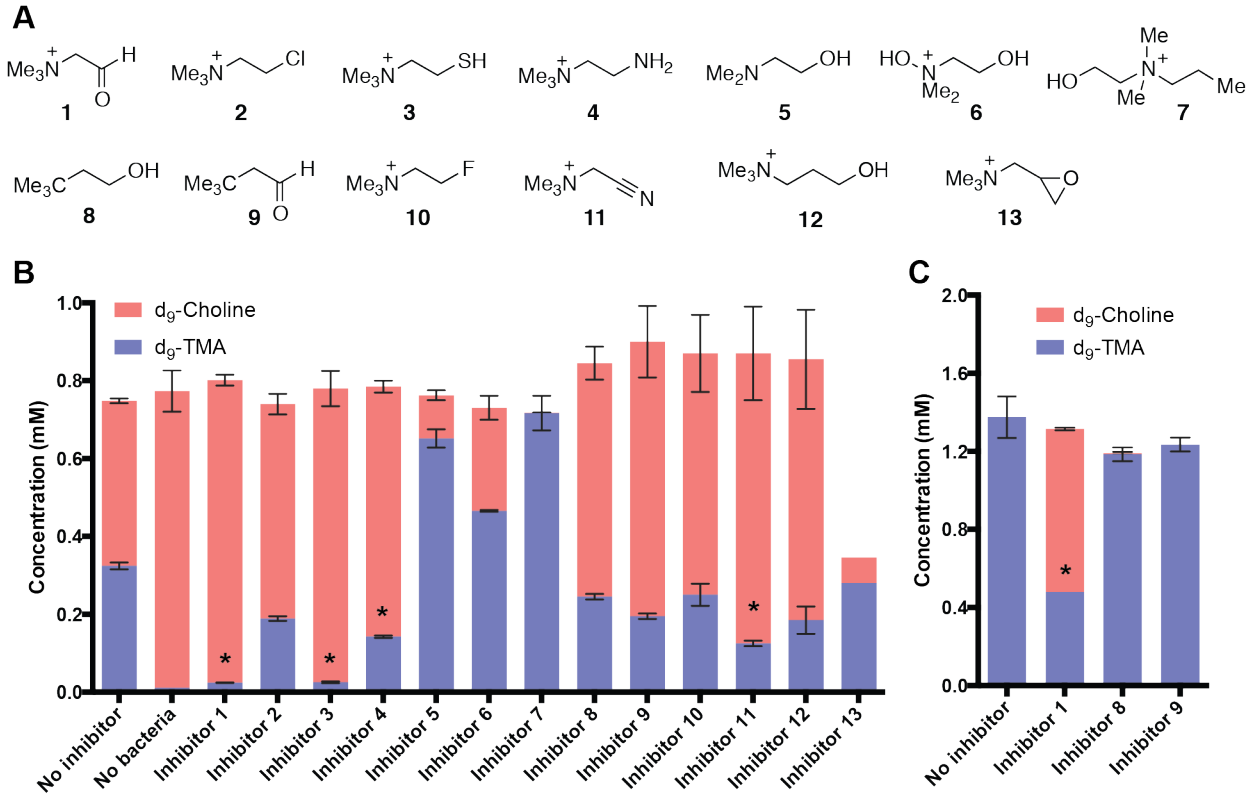


Figure 5.4: Small molecules inhibit anaerobic choline metabolism by *Escherichia coli* MS 200-1. **(A)** Rationally-designed molecules that were tested in our assays. **(B)** Results of an LC-MS/MS quantification of d_9-choline and d_9-TMA from cultures of *E. coli* MS 200-1, or **(C)** *Proteus mirabilis* ATCC 29906 grown for 3.5 h at 37 °C in BHI supplemented with 1 mM d_9-choline and 5 mM of each putative inhibitor. Stacked bar graphs represent the means ± standard deviations (SD) of three independent cultures; for compound **13**, only the results for one culture are shown. Inhibition greater than 50% is marked with *.

Closely examining the results presented in **Figure 5.4**, it is clear that the presence of a trimethylammonium group in a putative inhibitor is more likely to ensure effective inhibition of TMA formation, in line with the significant contribution of CutC active site residues interacting with this substituent to binding and catalysis (as reported in Chapter 4). In addition, the ability of a molecule to interact with Cys489 appears to correlate with its inhibitory activity. Both betaine aldehyde (compound **1**) and thiocholine (compound **3**) emerged as effective inhibitors of TMA production, drastically diminishing the levels of TMA (to less than 10% of levels without inhibitor) under the conditions tested (5 mM inhibitor). In addition, cyanocholine (compound **11**) appeared as a promising scaffold, lowering TMA to ~40% of levels without inhibitor. While we have thus identified small molecules that can modulate anaerobic choline catabolism by *E. coli*, there is significant room for improvements in potency and structural novelty to create better candidates for use within a host.

Our results with betaine aldehyde were in accordance with previous data from a study reporting that this molecule could inhibit TMA production (at 0.5 mM) from whole cells of *Streptococcus sanguis*, a *cut* gene cluster-harboring member of the oral cavity microbiota (23). Interestingly, betaine aldehyde is endogenous to our bodies, being produced as a short-lived intermediate in the oxidation of choline en route to glycine betaine in the mitochondria of liver and kidney cells (32). In order to determine the mechanism of inhibition by betaine aldehyde, future experiments should test whether this molecule can affect choline transport in bacteria (33), even it was not previously found to prevent low affinity choline transport in alveolar epithelial cells (34). Since choline metabolism in *E. coli* MS 200-1 (and most gut isolates (31)) is hypothesized to occur in a microcompartment, it is possible that betaine aldehyde could affect transport of choline into the microcompartment. This hypothesis could be tested in a similar manner as with whole cells, by using isolated microcompartments from a culture of *E. coli* MS 200-1 grown on choline as single carbon source (35). Another way in which betaine aldehyde could lower TMA production is by interfering with regulation of the *cut* pathway. This fact could be tested by comparing the transcription of proteins from the cluster when bacteria are grown on media containing choline with or without the addition of betaine aldehyde, using a procedure previously developed in the Balskus group

(31). Alternatively, the efficacy of betaine aldehyde against TMA production could be tested with bacteria grown to saturation and washed with buffer. Given that this molecule possesses an aldehyde moiety, we hypothesize that it could interfere with choline cleavage primarily by interacting with Cys489 of CutC to form a covalent bond.

The Hazen group at the Cleveland Clinic also recently conducted a study to discover small molecule inhibitors of TMA production by gut microbes (33). They tested several choline mimics (in which the N atom of choline was replaced with a C, Si or P atom, and glycyine betaine) and identified 3,3-dimethyl-1-butanol (DMB, compound **8**) as a weak inhibitor of in vivo TMA production by cells of *Proteus mirabilis* ATCC 29906 harboring the *cut* cluster (1 mM DMB elicited a 60% reduction in TMA levels). Moreover, DMB was found to significantly thwart the development of aortic root lesions in $ApoE^{-/-}$ mice that were fed a high choline diet. Metabolomics experiments indicated that DMB is rapidly metabolized in mice to 3,3-dimethylbutyric aldehyde (DMBA, compound **9**), a molecule that might also act as inhibitor of choline cleavage (33). In addition, it is likely that the observed phenotype in mice could be due to DMB or its metabolites interfering with other choline-binding proteins. In our experiments with *E. coli* MS 200-1, both DMB and DMBA appeared significantly less effective at preventing d_9-TMA formation, despite testing these molecules at concentrations 2.5-fold higher than those used by the Hazen group. We also incubated DMB or DMBA with cultures of the same strain of *Proteus mirabilis* that was used in their study, and we could not detect inhibition of choline cleavage (**Figure 5.4.C**). Even though our quantification of d_9-TMA for these experiments lacks some accuracy due to a conversion higher than the theoretical one, it is clear that neither DMB, nor DMBA, had any effect on TMA production from choline. In the experiments performed in the Hazen group, 8 mM DMB led to a ~80% reduction in TMA levels produced by cultures of *Proteus mirabilis* grown in rich media (nutrient broth) supplemented with inhibitor and 100 μM d_9-TMA (33). We believe that one reason behind these different results is the fact that experiments in our group were performed with cells grown under strict anaerobic conditions, whereas the Hazen group appeared to run experiments under aerobic or microaerophilic conditions (performing assays for TMA production in gas tight vials). This fact might explain why the conversion of d_9-choline

into d_9-TMA observed by the Hazen group was only ~8% with whole cells of *Proteus mirabilis*. Interestingly, in our experiments, betaine aldehyde retained its inhibitory activity with *Proteus mirabilis*, prompting us to further examine its specificity for CutC and its ability to prevent TMA formation by other *cut*-cluster containing bacteria.

5.2.2: Betaine aldehyde might function as a reversible covalent inhibitor of CutC

In order to determine how betaine aldehyde interacts with CutC, Michael Funk from the Drennan group at Massachusetts Institute of Technology soaked this compound into crystals of CutC at 50 mM concentration. The resulting crystal structure at 2.2 Å resolution (**Figure 5.5.A**) revealed the formation of a covalent, thiohemiacetal adduct with Cys489, which indicates that betaine aldehyde can directly interact with CutC and may function as a covalent, reversible inhibitor (**Figure 5.5.B**). It also appears that betaine aldehyde establishes similar interactions with other CutC active site residues as those observed in the choline-bound crystal structure and described in Chapter 4. Five CH–O interactions (3.3-3.7 Å) were observed between the methyl groups of the trimethylammonium moiety in betaine aldehyde and oxygen atoms of CutC active site residues (Asp216, Tyr208, Tyr506) or of the water molecule present in the CutC active site. Betaine aldehyde also interacts with Glu491 through a short hydrogen bonding interaction (2.4 Å) that might increase the stability of the thiohemiacetal formed with Cys489, by forcing the C–O bond of the intermediate out of the conformation needed for rehybridization. This fact is supported by the bond angles established between C1 and the surrounding atoms (~105-111 ^{o}) in the thiohemiacetal, indicating a ~sp^3 hybridization.

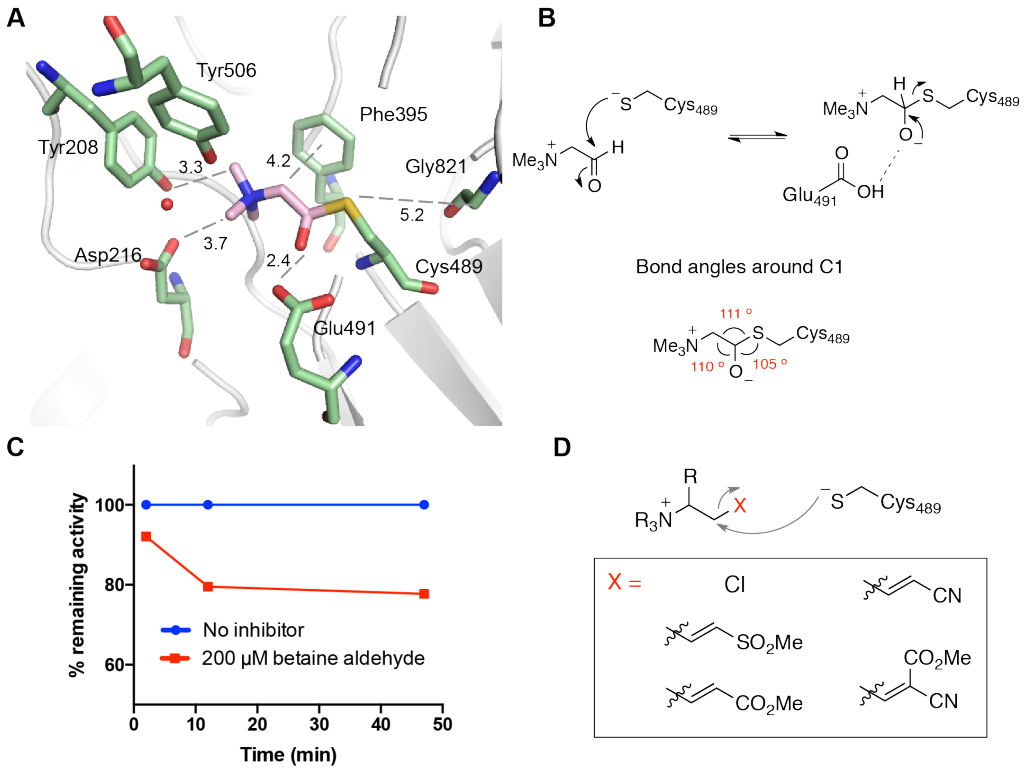


Figure 5.5: Betaine aldehyde can directly interact with CutC and may function as a reversible covalent inhibitor. **(A)** Crystal structure of CutC in complex with betaine aldehyde revealing a covalent bond between C1 of the inhibitor and S of Cys489. Active site residues are in green; betaine aldehyde in pink. **(B)** Mechanism for the formation of the thiohemiacetal between betaine aldehyde and Cys489. **(C)** Preliminary in vitro results for the inhibition of CutC activity by betaine aldehyde. CutC activity was measured using a coupled spectrophotometric assay with yeast alcohol dehydrogenase, as described in 'materials and methods'. The results of a representative experiment are shown. **(D)** Putative inhibitor scaffolds that could interact with Cys489. R = alkyl group.

In order to investigate the kinetic parameters of inhibition of CutC activity by betaine aldehyde, we performed a preliminary assay in which 10 µM of CutC dimer converted into the glycyl radical form was incubated with various concentrations of betaine aldehyde for different periods of time (**Figure 5.5.C**) and the ability of CutC to convert choline into acetaldehyde was measured via a spectrophotometric coupled assay with yeast alcohol dehydrogenase. Since the activity of CutC in the presence of betaine aldehyde is lower than without the addition of this molecule but remains constant over 30 min, these data support the proposed function of this molecule as a covalent reversible inhibitor (**Figure 5.5.C**). However, several experiments still need to be performed in order to fully classify the role of betaine aldehyde as inhibitor of CutC activity. Higher concentrations of betaine aldehyde would need to be tested in the assay described above to determine a dose-response curve and an IC50 value. In

addition, CutC should be incubated with betaine aldehyde for different amounts of time then dialyzed against storage buffer and tested in an activity assay, so as to establish whether the formation of the thiohemiacetal intermediate is reversible, and whether this molecule interacts preferentially with activated or unactivated CutC.

Considering the small molecular weight of betaine aldehyde, these results are very promising for the development of more complex and effective inhibitors that would take advantage of the catalytic cysteine residue present in the CutC active site (**Figure 5.5.D**) (36). Recent research aimed at identifying small molecule inhibitors of kinases showed that selectivity for a specific kinase can be achieved when a weakly active electrophilic moiety is added to an efficient ligand for that enzyme (37). In our survey of other structurally characterized choline-binding proteins, we could not identify any protein that contains a Cys residue near the active site (as described in Chapter 4). We thus believe that our rational design strategy is likely to yield novel, potent and specific CutC inhibitors.

5.2.3: Betaine aldehyde inhibits choline TMA-lyase activity in a panel of cut *gene cluster-containing bacteria*

In order to determine whether betaine aldehyde could inhibit TMA production by other bacteria harboring the *cut* gene cluster, we tested its efficacy at several different concentrations with cultures of Gram-positive (*Clostridium sporogenes* ATCC 15579) and Gram-negative (*Proteus mirabilis* BB2000*, Escherichia coli* MS 69-1*, Klebsiella sp.* MS 93-2) human gut isolates, using similar experimental conditions as with *E. coli* MS 200-1. As described in Chapter 2, we classified *cut* clusters into two groups based on their gene content: type I was present primarily in *Firmicutes* and contained ~20 genes, while type II was present in *Proteobacteria* and contained only ~10 genes (31). We do not know yet whether the protein microcompartments from the two types of organisms are different, since type I clusters contain more putative microcompartment genes than type II clusters. Nonetheless, **Figure 5.6.A** reveals that betaine aldehyde is effective against all the strains assayed, including the Firmicute *Clostridium sporogenes*, an encouraging result given that members of the Firmicutes phylum constitute ~50% of gut bacteria (38). The IC_{50} values we observed are within 7-fold of each other, perhaps due to differences in

cell wall or microcompartment architecture and permeability, or the activity of efflux pumps, known to contribute to drug resistance particularly in Gram-positive bacteria (39). In addition, one of the caveats of this experiment is the fact that the inhibitor is present in the medium as the cells are growing. Thus, the efficacy of betaine aldehyde should also be tested with bacteria grown to saturation in the presence of choline (to induce the transcription of the cut cluster) and washed with buffer to ensure that bacterial growth over the duration of the experiment is minimal. Taken together, our data indicate that small molecules can interfere with anaerobic choline metabolism in gut microbes, regardless of the type of *cut* cluster or cell wall and membrane architecture.

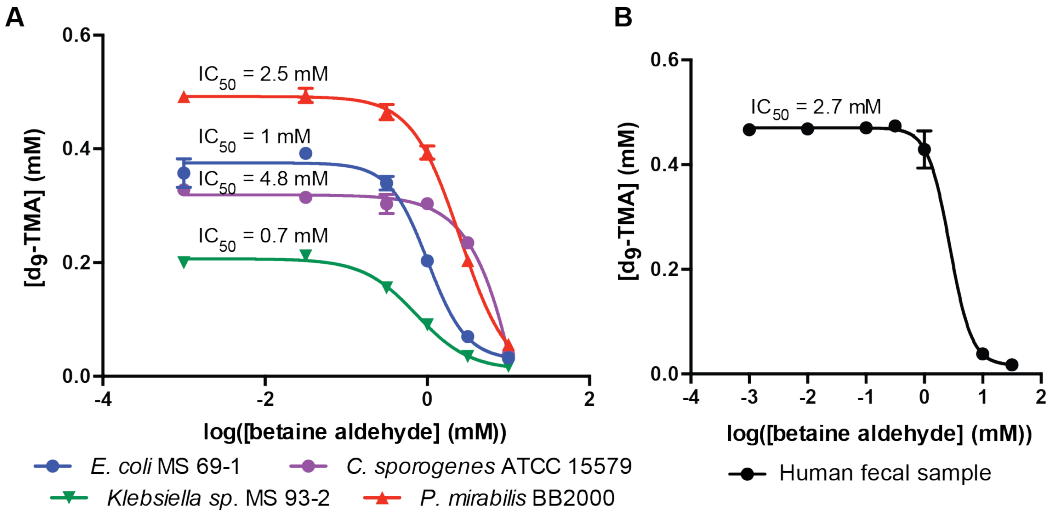


Figure 5.6: Betaine aldehyde inhibits d_9-TMA formation by **(A)** a panel of *cut* cluster-harboring bacteria (*Escherichia coli* MS 69-1, *Klebsiella sp*. MS 93-2, *Proteus mirabilis* BB2000, *Clostridium sporogenes* ATCC 15579) and **(B)** by a human fecal sample in a dose-dependent manner. Production of d_9-TMA was measured by LC-MS/MS and each experiment was performed in triplicate for (A) and duplicate for (B).

5.2.4: Betaine aldehyde inhibits TMA generation by a human fecal sample

Our previous experiments detailed in Chapter 2 have shown that the *cut* pathway likely represents a major mechanism for TMA production from choline by gut microbes (31). One method to verify this hypothesis is to selectively inhibit CutC from a consortium of gut bacteria (such as the one found in a human fecal sample) using a small molecule inhibitor and assess the community's ability to convert choline into TMA. In our earlier studies of the *cut* gene cluster (Chapter 4), we identified ex vivo conditions that allowed us to observe d_9-TMA production when incubating fecal samples with d_9-choline

(31). Employing these conditions and adding various amounts of betaine aldehyde at the beginning of incubation, we observed that this molecule is capable of abolishing TMA formation when used at mM concentrations (**Figure 5.6.B**). Even though we do not have yet a complete understanding of the proteins betaine aldehyde is interacting with, this result suggests that the efficacy of betaine aldehyde extends beyond the bacterial strains that we had tested to an intact community containing a diverse range of organisms. In addition, these data complement studies with gnotobiotic mice performed by the Rey group (40) (mentioned in Chapter 2), thus confirming that CutC-mediated TMA production is likely the major gut bacterial pathway for choline conversion into TMA.

5.2.5: Development of a growth-based assay for high throughput screening for inhibitors of choline cleavage by Escherichia coli *MS 69-1*

Our second, orthogonal approach towards discovering small molecule inhibitors of anaerobic microbial choline metabolism involves a high throughput screen that takes advantage of the ability of *cut* gene cluster harboring bacteria to utilize choline as a sole carbon source. Dr. Ana Martínez-del Campo has recently shown that organisms such as *Escherichia coli* MS 200-1 cannot grow in media containing choline as single carbon source when *cutC* or other genes from the *cut* gene cluster have been deleted (manuscript in preparation). However, these mutations do not impact growth in rich media or when alternate carbon sources, such as glucose or glycerol, are present. Our aim is thus to identify small molecules that inhibit growth on choline, but do not affect bacterial growth on an alternate carbon source, such as glycerol or glucose, so as to differentiate inhibitors of choline metabolism from molecules that possess antibiotic properties (**Figure 5.7.A**). To this end, we will utilize two strains of bacteria that harbor the *cut* gene cluster but differ in the cell membrane and possibly in the architecture of their *cut* microcompartment: *Escherichia coli* MS 69-1 (a Gram-negative strain with a type II *cut* gene cluster) and *Lactobacillus paralimentarius* (a Gram-positive strain with a type I *cut* gene cluster). We anticipate that using these two bacterial strains in the high throughput screen side-by-side will enable us to discover broad-spectrum inhibitors of microbial TMA production.

The success of the high throughput screen depends primarily on our ability to identify minimal media formulations for each strain of bacteria such that growth is dependent solely on the utilization of the carbon source provided. Our starting point for developing a minimal media for *E. coli* MS 69-1 was the "no carbon E" media utilized for achieving ethanolamine-dependent growth of *Salmonella enterica* (41). Marina Orman and I optimized the amounts of vitamins, minerals, electron acceptor (sodium fumarate) and nitrogen source (cas amino acids) such that a maximum OD_{600} value of 0.4 could be achieved with the addition of 30 mM choline, and an OD_{600} value of 1.2 with 30 mM glycerol (**Figure 5.7.B**). In our initial experiments, we observed a significant (~8 h) delay in growth on choline. Previous experiments in our group determined that the *cut* cluster is induced many fold in the presence of choline (31), so we hypothesized that this delay could be due to a slow induction of the *cut* genes. As a result, we supplemented the overnight culture used for the inoculum with 1 mM choline, and were able to diminish the delay in growth. Unfortunately, using the same strategy for growing *E. coli* MS 69-1 with glycerol as carbon source did not reduce the growth delay observed initially and ongoing experiments are aimed at addressing this issue. In addition, future efforts in the Balskus group will focus on optimizing a minimal media formulation for the growth of *L. paralimentarius* on glucose or choline as sole carbon source.

Since betaine aldehyde was effective at preventing anaerobic choline utilization when *E. coli* MS 69-1 was grown in BHI supplemented with d_9-choline (**Figure 5.6.A**), we tested whether it could interfere with choline utilization in the growth-based assay for this bacterial strain carried out in a 96-well plate. We observed that a high concentration of betaine aldehyde (20 mM) effectively abolished growth on choline but had no impact on the ability of *E. coli* MS 69-1 to utilize glycerol as carbon source (**Figure 5.7.B**). It is thus clear that betaine aldehyde interferes with anaerobic choline utilization, although further experiments will need to be done to elucidate the protein target(s) of this molecule (proteins from the *cut* gene cluster or other choline-binding proteins). Nonetheless, we are able to use betaine aldehyde as positive control for the high throughput assay and to calculate a Z' factor of this assay (a statistical parameter for assay quality assessment (42)), based on the OD_{600} of cultures grown on choline with or without the addition of betaine aldehyde. The value we obtained for the Z' factor of this assay is ~0.84 at

14 h of incubation (out of a maximum Z' of 1), which indicates that our assay is robust enough to be used as part of the high throughput screen.

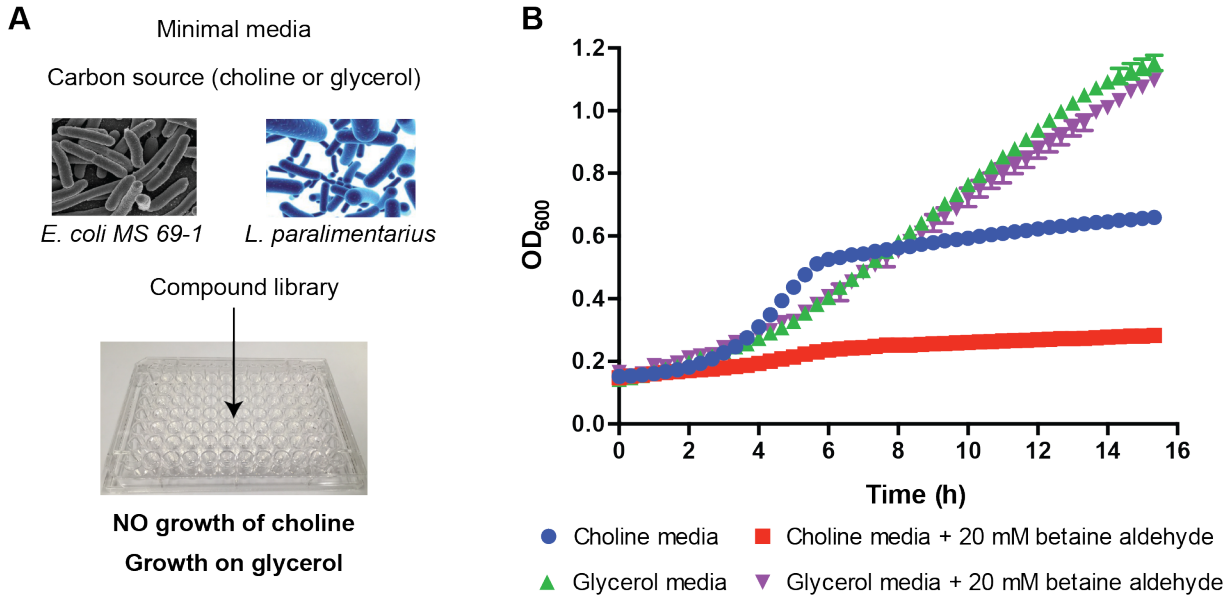


Figure 5.7: Growth-based assay for high throughput screening to identify inhibitors of anaerobic choline metabolism. **(A)** Differential growth-based screening approach with *Escherichia coli* MS 69-1 and *Lactobacillus paralimentarius*. **(B)** Representative growth curves for *E. coli* MS 69-1 in the presence and absence of CutC inhibitor betaine aldehyde.

Our future work will focus on validation of this screening strategy and identification of initial lead molecules. To this end, we will utilize compound libraries from the Institute of Chemistry and Cell Biology at Harvard Medical School. We will follow up any hits with assays for human cell cytotoxicity (43) and choline transport inhibition (33). We will also evaluate each hit's efficacy for inhibiting TMA production in both fecal samples and *cut* gene cluster-harboring gut bacteria. Since multiple *cut* genes are essential for growth on choline as sole carbon source, follow-up experiments will be performed to determine the mechanism by which our lead molecules interfere with anaerobic choline utilization. For instance, their ability to inhibit CutC, CutD, CutO or CutF in vitro could be determined using assays already developed in the Balskus group (31, 44). Electron microscopy might reveal whether inhibitors affect the assembly of the microcompartment (45), and interactions with other proteins could be elucidated by affinity chromatography using inhibitors immobilized on a solid phase (46). We expect that

further optimization of initial leads will afford potent molecules that could eventually be utilized with different mouse models of liver and heart disease in order to assess the impact of choline cleavage on host metabolism and disease progression.

5.2.6: Conclusions

In summary, we have demonstrated that small molecules can act as inhibitors of anaerobic choline metabolism in both Gram-positive and Gram-negative bacteria, as well as in human fecal samples. Using both structure-guided rational design and high throughput screening, we are now poised to discover improved inhibitors that could function as tools for the study of choline metabolism in complex microbial communities and in the host, or could be further optimized into novel therapeutics. Disorders involving gut microbial metabolism are currently treated with antibiotics, which damage this microbial community, abolishing its beneficial functions, and occasionally allowing multi-drug resistant pathogens to thrive (47). Because impairing choline metabolism does not appear to impact bacterial viability, the small molecules we develop may have significantly improved safety profiles. Overall, targeting specific gut microbial functions with small molecules is a new approach for studying gut microbial communities that may be applied broadly to other metabolic pathways.

5.3: Materials and methods

5.3.1: Materials and general methods

All chemicals were purchased from Sigma-Aldrich, except for d_9-trimethylamine (TMA) hydrochloride (CDN Isotopes) and (trimethyl-d_9)-choline chloride (Cambridge Isotope Laboratories). Acetonitrile and water used for liquid chromatography-mass spectrometry (LC-MS) were B&J Brand high-purity solvents (Honeywell Burdick & Jackson). All NMR solvents were purchased from Cambridge Isotope Laboratories. LC-MS/MS analysis for d_9-TMA and d_9-choline was performed without derivatization, as described in Chapter 2, in our research labs in the Department of Chemistry and Chemical Biology, Harvard University, on an Agilent 6410 Triple Quadrupole LC/MS instrument (Agilent Technologies, Wilmington, DE). The analysis for d_9-choline was performed under very similar

conditions to that for d_9-TMA, by modifying the precursor-product ion pair (m/z 113.2→m/z 45.1) and the collision energy (21 V).

For LC-MS/MS analyses, bacteria were grown on a 200 μL scale in 96-well flat-bottom plates (Falcon) sealed with aluminum foil (VWR). Optimization of high throughput screening media formulation and growth assays were performed using either the same type of 96-well plates or in 384-well clear plates (Corning); lids were kept on the plate during the incubation but the sides were sealed with electric tape to prevent the media from evaporating. All growth and kinetic assays were conducted using a PowerWave HT Microplate Spectrophotometer (BioTek). Growth assays were performed at 37 °C, and the absorbance of each well was monitored at 600 nm for 16 h, with reads 20 min apart. Kinetic assays were performed at 22 °C, the absorbance of each well was monitored at 340 nm for 10 min, reads being 20 s apart. Path length correction was employed such that absorbance values were reported for a path length of 1 cm. Data were analyzed using Gen5 software and EC_{50} values were calculated using GraphPad Prism.

Protein purification was performed for CutC (–52) aa wild-type as described in Chapter 4. Protein concentrations were determined according to the method of Bradford (48) using bovine serum albumin (BSA) as a standard or using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific) for CutC (ε = 128,870 M^{-1} cm^{-1}). The extinction coefficient for CutC (–52) aa was obtained using the ExPASy protparam tool (http://www.expasy.org/tools/protparam.html). All anaerobic experiments were conducted in an MBraun glovebox (MBraun) under an atmosphere consisting of 99.997% N_2 with less than 5 ppm O_2, or in an anaerobic chamber (Coy Laboratory Products) under an atmosphere of 98% N_2 and 2% H_2.

Proton nuclear magnetic resonance (^{1}H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded in the Magnetic Resonance Laboratory in the Harvard University Department of Chemistry and Chemical Biology on a Varian Inova-500 (500 MHz, 126 MHz) NMR spectrometer. Chemical shifts are reported in parts per million downfield from tetramethylsilane using the solvent resonance as an internal standard for ^{1}H (CD_3OD, δ_H = 3.31 ppm) and ^{13}C (CD_3OD, δ_C = 49 ppm). Data are reported as follows: chemical shift, integration multiplicity (s = singlet, d = doublet, t = triplet),

coupling constant, integration, and assignment. NMR spectra were visualized using iNMR Version 5.4.5 (Mestrelab Research).

High-resolution mass spectral (HRMS) data for the synthetic compounds were obtained in the Small Molecule Mass Spectrometry Facility, FAS Division of Science at Harvard University on an Agilent 6210 TOF fitted with a dual-spray electrospray ionization (ESI) source. The capillary voltage was set to 4.5 kV and the end plate offset to –500 V, while the drying gas temperature was maintained at 190 ^{o}C with a flow rate of 8 L/min and a nebulizer pressure of 21.8 psi. Liquid chromatography (LC) was performed using an Agilent Technologies 1100 series LC with 50% H_2O and 50% acetonitrile as solvent.

5.3.2: Crystallization and structure determination of wild-type CutC bound to betaine aldehyde

Crystallization conditions for CutC with a 52-amino acid N-terminal truncation and His_6-tag were identified as previously described in Chapter 4. Crystals were obtained in 16-19% PEG 8000, 0.2-0.4 M LiCl, and 0.1 M Tris 8.0 by hanging drop vapor diffusion at 21 ºC. A solution of CutC in 50 mM potassium phosphate pH 8.0, 50 mM KCl, and 10% (v/v) glycerol was preincubated at 80 µM with 10 mM betaine aldehyde. Protein and precipitant were mixed in a 1:1 ratio and incubated for 3-5 days. Crystals were cryoprotected in a stabilization solution containing 20% (v/v) glycerol, 25% (w/v) PEG 8000, 0.5 M lithium chloride, and 0.1 M Tris pH 8.0 and supplemented with 50 mM betaine aldehyde.

Diffraction images were collected at the Advanced Photon Source beamline 24ID-C at a wavelength of 0.9795 Å on a Pilatus 6M detector (Dectris). The diffraction patterns were indexed, integrated, and scaled in HKL2000 (49). Crystals diffracted to 2.20-Å resolution in space group $P2_1$ with cell edges a = 105.2 Å, b = 234.7 Å, c = 159.0 Å, β = 109.0º. The structure was solved by molecular replacement in the Phenix implementation of Phaser (50) using the wild type monomer with all ligands removed as the search model. Betaine aldehyde covalently bound to Cys489 was modeled into difference density with Cys489 removed. The *S* configuration is probable based on the expected binding mode prior to nucleophilic attack; however, the stereochemistry cannot be unambiguously assigned by the electron density alone. Parameter files were made in Phenix eLBOW (51). Initial refinement was performed in

phenix.refine (51) with manual rebuilding in Coot (52). Water molecules were placed automatically. The final model contains all residues found in the wild type structure.

5.3.3: In vitro assay for inhibition of CutC activity by betaine aldehyde

CutC (10 μM dimer) was first activated to the glycyl radical form in an MBraun glove box in a 1.5 mL polypropylene Eppendorf tube. Activation assays contained 50 mM potassium phosphate pH 8, 50 mM potassium chloride, 150 μM sodium dithionite, 200 μM SAM, 10 μM CutC dimer, and 40 μM CutD in a total volume of 450 μL. A mixture of buffer, CutD and sodium dithionite was incubated for 20 min prior to addition of SAM and CutC, then glycyl radical formation was carried out for 1 h at room temperature. The assay was divided into 5 tubes, to which betaine aldehyde was added in order to reach final concentrations of 5-200 μM. The mixtures were incubated for 2, 12 or 47 min prior to a 3,333-fold dilution for the kinetic activity assay. A spectrophotometric coupled assay monitoring oxidation of NADH by yeast alcohol dehydrogenase (YADH) was used to determine the kinetics of acetaldehyde production and thereby choline cleavage**.** All assays contained 200 μM NADH, 50 mM potassium phosphate pH 8, 50 mM potassium chloride, diluted activation mixture, YADH (0.4 μM) and choline (2 mM), in a total volume of 200 μL. The assay was carried out in a 96-well plate and the NADH absorbance at 340 nm was monitored for 5 min. The absorbance decreased linearly, indicating a constant concentration of activated enzyme during the course of our measurements.

5.3.4: LC-MS/MS assay for determining choline TMA-lyase activity of cut *gene cluster-harboring bacteria*

A solution of inhibitor in 50 mM potassium phosphate buffer pH 7.4 was added to 200 μL BHI supplemented with 1 mM d_9-choline to reach the final concentrations indicated in **Figures 5.4** and **5.6**. The mixture was inoculated with 2% v/v of an overnight culture (of *Escherichia coli* MS 200-1, *Escherichia coli* MS 69-1, *Proteus mirabilis* ATCC 29906, *Klebsiella sp*. MS 93-2, *Proteus mirabilis* BB2000, *Clostridium sporogenes* ATCC 15579) that had been grown to saturation in BHI with 1 mM d_9-choline. For *C. sporogenes* ATCC 15579, the media used instead of BHI was Reinforced Clostridial Media (Becton Dickinson). All cultures were incubated at 37 ^{o}C in 96-well plates with aluminum seals,

inside a vinyl anaerobic chamber (Coy Laboratories). Incubation times varied as follows: 3.5 h for *E. coli* MS 200-1 and *P. mirabilis* ATCC 29906, 8 h for *C. sporogenes* ATCC 15579, and 4 h for the other cultures. An aliquot of each culture (8 µL) was diluted (up to 2500-fold) with a mixture of 80% acetonitrile, 20% 5 mM ammonium formate buffer pH 4, and analyzed by LC-MS/MS for d_9-TMA and d_9-choline.

5.3.5: Ex vivo incubation for inhibition of choline TMA-lyase activity in a fecal sample

A frozen fecal sample from a healthy individual was diluted 1:10 (wt/v) in sterile anaerobic PBS. The dilution was vortexed for 1 min and allowed to settle for 15 min at room temperature inside a vinyl anaerobic chamber. A 95 µL aliquot of the resulting supernatant was transferred to a 96-well plate containing 5 µL of different concentrations of betaine aldehyde (10 µM–31 mM final concentration), and the mixtures were incubated for 16 h at 37°C. An aliquot of each culture (8 µL) was diluted 2500-fold with a mixture of 80% acetonitrile, 20% 5 mM ammonium formate buffer pH 4, and analyzed by LC-MS/MS for d_9-TMA and d_9-choline.

5.3.6: Growth-based assay for high throughput screening with E. coli *MS 69-1*

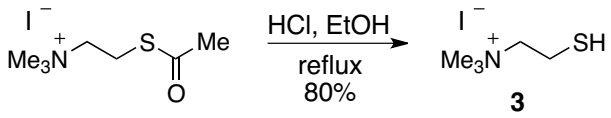
The optimized minimal media for growth of *E. coli* MS 69-1 contained: 29 mM potassium phosphate monobasic, 28.5 mM potassium phosphate dibasic, 16.7 mM ammonium sodium phosphate dibasic, 1 mM magnesium sulfate, 1% v/v ATCC vitamin supplement, 1% v/v ATCC trace mineral supplement, 30 mM choline or 30 mM glycerol, 40 mM sodium fumarate dibasic, 0.4% wt/v casamino acids (Becton Dickinson). For the growth inhibition assay, betaine aldehyde was added to media in a 96-well plate for a final concentration of 20 mM, and the mixture was inoculated with 1% v/v of an overnight culture grown in BHI supplemented with 1 mM d_9-choline, in a total volume of 200 µL. The cultures were incubated at 37 ^{o}C for 16 h in an Mbraun glove box under nitrogen atmosphere and OD_{600} values were measured using a BioTek spectrophotometer, as mentioned in 'materials and general methods'. The Z' value for the assay was calculated using OD_{600} values at 14 hours for cultures without inhibitor (negative control) and with inhibitor (positive control) according to the following formula: Z' = 1 –

$[(3\sigma_{neg} + 3\sigma_{pos})/(|\mu_{neg} - \mu_{pos}|)]$, where μ is the mean absorbance of three cultures, and σ is the standard deviation.

5.3.7: Chemical synthesis procedures and characterization data

Betaine aldehyde chloride (compound **1**) was prepared according to the procedure described in Chapter 3 (44).

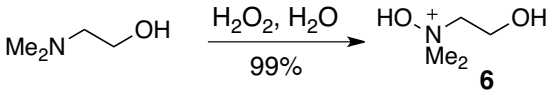
Synthesis of 2-mercapto-*N*,*N*,*N*-trimethylethanaminium iodide (compound **3**)



3

A solution of acetylthiocholine iodide (390 mg, 1.35 mmol) in ethanol (19 mL) was sparged with Ar in an oven-dried 100 mL round bottom flask for 15 min. Upon addition of concentrated (12 M) HCl (0.7 mL), the flask was fitted with a reflux condenser and the mixture was heated at 37 ^{o}C for 16 h under an Ar atmosphere. Upon cooling, the mixture was transferred to an Ar-flushed flask via syringe, and most solvent was removed *in vacuo*. The resulting gel was dissolved in anaerobic water, frozen in liquid N_2, and lyophilized overnight to afford a white solid (291 mg, 1.07 mmol, 80% yield). HRMS (ESI): calc'd for $C_5H_{14}NS^+$ $[M]^+$, 120.0847; found, 120.0851. ^{1}H-NMR (500 MHz; CD_3OD) δ: 3.56 (t, J = 8.4 Hz, 2H, C**H**$_2$SH), 3.17 (s, 9H, NC**H**$_3$), 2.94 (t, J = 8.3 Hz, 2H, NC**H**$_2$). ^{13}C NMR (126 MHz; CD_3OD) δ: 69.5, 53.69, 53.66, 53.63, 17.8.

Synthesis of *N*-(2-hydroxyethyl)-*N*,*N*-dimethylhydroxylammonium (compound **6**)

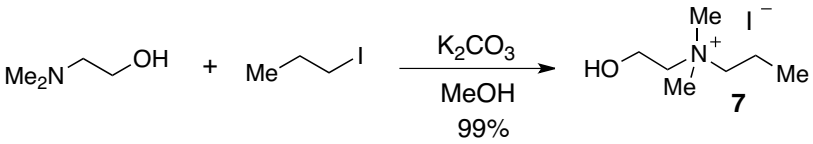


6

To a solution of 2-dimethylaminoethanol (93 mg, 1.04 mmol) in water (5 mL), a solution of hydrogen peroxide in water (30% v/v, 0.24 mL, 2.09 mmol) was added. The reaction mixture was allowed to stir at room temperature overnight, then the solvent and excess reagent were removed *in vacuo* to afford a colorless gel (108 mg, 1.03 mmol, 99% yield). HRMS (ESI): calc'd for $C_4H_{12}NO_2^+$ $[M]^+$,

106.0868; found, 106.0873. ^{1}H-NMR (500 MHz; CD_3OD) δ: 4.04-4.02 (m, 2H, C**H**$_2$OH), 3.49-3.47 (m, 2H, NC**H**$_2$), 3.27 (s, 6H, NC**H**$_3$). ^{13}C NMR (126 MHz; CD_3OD) δ: 72.7, 59.2, 57.4.

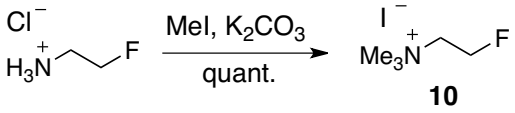
Synthesis of *N*-(2-hydroxyethyl)-*N*,*N*-dimethylpropan-1-aminium iodide (compound **7**)



7

To a solution of 2-dimethylaminoethanol (200 mg, 2.03 mmol) in methanol (5 mL), potassium carbonate (364 mg, 2.64 mmol) and 1-iodopropane (690 mg, 396 µL, 4.06 mmol) were added, and the mixture was stirred overnight at room temperature, under a nitrogen atmosphere. The mixture was filtered, and the solvent was removed *in vacuo*. The resulting solid was dissolved in water and washed twice with diethyl ether and twice with ethyl acetate. The water was removed *in vacuo* to afford a white solid (333 mg, 1.29 mmol, 63% yield). HRMS (ESI): calc'd for $C_7H_{18}NO^+$ $[M]^+$, 132.1388; found, 132.1393. ^{1}H-NMR (500 MHz; CD_3OD) δ: 4.01 (m, 2H, C**H**$_2$OH), 3.48 (t, *J* = 4.9 Hz, 2H, C**H**$_2$CH$_2$OH), 3.37 (dt, *J* = 8.2, 4.3 Hz, 2H, C**H**$_2$N), 3.16 (s, 6H, C**H**$_3$N), 1.87-1.79 (m, 2H, C**H**$_2$-CH$_3$), 1.01 (t, *J* = 7.3 Hz, 3H, C**H**$_3$CH$_2$). ^{13}C NMR (126 MHz; CD_3OD) δ: 68.1, 66.6, 56.7, 52.41, 52.35, 17.1, 10.9.

Synthesis of 2-fluoro-*N*,*N*,*N*-trimethylethanaminium iodide (compound **10**)



10

To a solution of 2-fluoroethylamine hydrochloride (100 mg, 0.9 mmol) in methanol (4.5 mL), potassium carbonate (481 mg, 3.48 mmol) and iodomethane (642 mg, 280 µL, 4.52 mmol) were added, and the mixture was stirred overnight at room temperature, under a nitrogen atmosphere. The mixture was filtered, and the solvent was removed *in vacuo*. The resulting white solid was washed four times with acetone (15 mL), and dried *in vacuo*, but potassium salts still remained as impurities (490.6 mg, quantitative yield by NMR, 43% purity calculated based on weight). HRMS (ESI): calc'd for $C_5H_{13}FN^+$ $[M]^+$, 106.1032; found, 106.1034. ^{1}H-NMR (500 MHz; CD_3OD) δ: 5.04-4.92 (m, 2H, C**H**$_2$F), 3.87-3.80 (m, 2H, NC**H**$_2$), 3.27 (s, 9H, NC**H**$_3$). ^{13}C NMR (126 MHz; CD_3OD) δ: 79.4, 54.9, 49.8.

5.4: References

1. Qin J*, et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59-65.

2. Holmes E, Li JV, Athanasiou T, Ashrafian H, & Nicholson JK (2011) Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol* 19(7):349-359.

3. Nicholson JK*, et al.* (2012) Host-gut microbiota metabolic interactions. *Science* 336(6086):1262-1267.

4. Wallace BD*, et al.* (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* 330(6005):831-835.

5. Wallace BD & Redinbo MR (2013) The human microbiome is a source of therapeutic drug targets. *Curr Opin Chem Biol* 17(3):379-384.

6. Sherriff JL, O'Sullivan TA, Properzi C, Oddo JL, & Adams LA (2016) Choline, its potential role in nonalcoholic fatty liver disease, and the case for human and bacterial genes. *Adv Nutr* 7(1):5-13.

7. Gaci N, Borrel G, Tottey W, O'Toole PW, & Brugere JF (2014) Archaea and the human gut: new beginning of an old story. *World J Gastroentero* 20(43):16062-16078.

8. Krueger SK & Williams DE (2005) Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol Therapeut* 106(3):357-387.

9. al-Waiz M, Mikov M, Mitchell SC, & Smith RL (1992) The exogenous origin of trimethylamine in the mouse. *Metabolism* 41(2):135-136.

10. Brown JM & Hazen SL (2015) The gut microbial endocrine organ: bacterially derived signals driving cardiometabolic diseases. *Annu Rev Med* 66:343-359.

11. Tang WH*, et al.* (2013) Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 368(17):1575-1584.

12. Koeth RA*, et al.* (2013) Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* 19(5):576-585.

13. Wang Z*, et al.* (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57-63.

14. Seldin MM*, et al.* (2016) Trimethylamine N-oxide promotes vascular inflammation through signaling of mitogen-activated protein kinase and nuclear factor-κB. *J Am Heart Assoc* 5(2):e002767.

15. Zhu W*, et al.* (2016) Gut Microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* 165(1):111-124.

16. Collins HL*, et al.* (2016) L-Carnitine intake and high trimethylamine N-oxide plasma levels correlate with low aortic lesions in ApoE(-/-) transgenic mice expressing CETP. *Atherosclerosis* 244:29-37.

17. Shih DM*, et al.* (2015) Flavin containing monooxygenase 3 exerts broad effects on glucose and lipid metabolism and atherosclerosis. *J Lipid Res* 56(1):22-37.

18. Channon HJ, Hanson SWF, & Loizides PA (1942) The effect of variations of diet fat on dietary fatty livers in rats. *Biochem J* 36:214-220.

19. Dumas ME*, et al.* (2006) Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci USA* 103(33):12511-12516.

20. Le Roy T*, et al.* (2013) Intestinal microbiota determines development of non-alcoholic fatty liver disease in mice. *Gut* 62(12):1787-1794.

21. Dowman JK, Armstrong MJ, Tomlinson JW, & Newsome PN (2011) Current therapeutic strategies in non-alcoholic fatty liver disease. *Diabetes Obes Metab* 13(8):692-702.

22. Messenger J, Clark S, Massick S, & Bechtel M (2013) A review of trimethylaminuria: (fish odor syndrome). *J Clin Aesthet Dermatol* 6(11):45-48.

23. Chao CK & Zeisel SH (1990) Formation of trimethylamine from dietary choline by *Streptococcus sanguis* I, which colonizes the mouth. *J Nutr Biochem* 1(2):89-97.

24. Hayward HR & Stadtman TC (1960) Anaerobic degradation of choline. II. Preparation and properties of cell-free extracts of *Vibrio cholinicus*. *J Biol Chem* 235:538–543.

25. Hayward HR (1960) Anaerobic degradation of choline. III. Acetaldehyde as an intermediate in the fermentation of choline by extracts of *Vibrio cholinicus*. *J Biol Chem* 235(12):3592-3596.

26. Bradbeer C (1965) Clostridial fermentations of choline and ethanolamine. I. Preparation and properties of cell-free extracts. *J Biol Chem* 240(12):4669-4674.

27. Sandhu SS & Chase T (1986) Aerobic degradation of choline by Proteus mirabilis: enzymatic requirements and pathway. *Can J Microbiol* 32(9):743-750.

28. Canny GO & McCormick BA (2008) Bacteria in the intestine, helpful residents or enemies from within? *Infect Immun* 76(8):3360-3373.

29. Koeth RA*, et al.* (2014) γ-Butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. *Cell Metab* 20(5):799-812.

30. Craciun S & Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109(52):21307-21312.

31. Martinez-del Campo A*, et al.* (2015) Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *MBio* 6(2):e00042-15.

32. Craig SAS (2004) Betaine in human nutrition. *Am J Clin Nutr* 80(3):539-549.

33. Wang ZN*, et al.* (2015) Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell* 163(7):1585-1595.

34. Dodia C, Fisher AB, Chander A, & Kleinzeller A (1992) Inhibitors of choline transport in alveolar type II epithelial cells. *Am J Resp Cell Mol* 6(4):426-429.

35. Yeates TO, Crowley CS, & Tanaka S (2010) Bacterial microcompartment organelles: protein shell structure and evolution. *Annu Rev Bioph* 39:185-205.

36. Hopkins AL, Keseru GM, Leeson PD, Rees DC, & Reynolds CH (2014) The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* 13(2):105-121.

37. Liu QS*, et al.* (2013) Developing irreversible inhibitors of the protein kinase cysteinome. *Chem Biol* 20(2):146-159.

38. Donaldson GP, Lee SM, & Mazmanian SK (2016) Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* 14(1):20-32.

39. Sun JJ, Deng ZQ, & Yan AX (2014) Bacterial multidrug efflux pumps: Mechanisms, physiology and pharmacological exploitations. *Biochem Bioph Res Co* 453(2):254-267.

40. Romano KA, Vivas EI, Amador-Noguez D, & Rey FE (2015) Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. *MBio* 6(2):e02481-14.

41. Price-Carter M, Tingey J, Bobik TA, & Roth JR (2001) The alternative electron acceptor tetrathionate supports B_{12}-dependent anaerobic growth of *Salmonella enterica serovar typhimurium* on ethanolamine or 1,2-propanediol. *J Bacteriol* 183(8):2463-2475.

42. Zhang JH, Chung TDY, & Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 4(2):67-73.

43. Miret S, De Groene EM, & Klaffke W (2006) Comparison of in vitro assays of cellular toxicity in the human hepatic cell line HepG2. *J Biomol Screen* 11(2):184-193.

44. Craciun S, Marks JA, & Balskus EP (2014) Characterization of choline trimethylamine-lyase expands the chemistry of glycyl radical enzymes. *ACS Chem Biol* 9(7):1408-1413.

45. Jameson E*, et al.* (2015) Anaerobic choline metabolism in microcompartments promotes growth and swarming of *Proteus mirabilis*. *Environ Microbiol*.

46. Ziegler S, Pries V, Hedberg C, & Waldmann H (2013) Target Identification for Small Bioactive Molecules: Finding the Needle in the Haystack. *Angew Chem Intl Ed* 52(10):2744-2792.

47. Macfarlane S (2014) Antibiotic treatments and microbes in the gut. *Environ Microbiol* 16(4):919-924.

48. Stoscheck CM (1990) Quantitation of protein. *Method Enzymol* 182:50-68.

49. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.

50. Mccoy AJ*, et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.

51. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66:213-221.

52. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D* 66:486-501.